****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 26 16:16:03 2000; MasPar time 9.87 Seconds
731.891 Million cell updates/sec
Tabular output not generated.

Title: >US-09-308-435-2
Description: (28-260) from US09308435.pep
Perfect Score: 1594
Sequence: 1 CSPHIIETNEVALKLNYHPA.....QKNLESYQKDAKELKGKRNR 233

Scoring table: PAM 150
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 48.589; Variance 103.194; scale 0.471

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1594	100.0	260	1	HPA3_HELPY	NEURAMINYLLACTOSE-BIND	1.03e-279
2	1580	99.1	260	1	HPA0_HELPY	NEURAMINYLLACTOSE-BIND	6.22e-277
3	1578	99.0	260	1	HPA2_HELPY	NEURAMINYLLACTOSE-BIND	1.55e-276
4	1551	97.3	260	1	HPA1_HELPY	NEURAMINYLLACTOSE-BIND	3.53e-271
5	848	53.2	125	1	HPAA_HELNE	NEURAMINYLLACTOSE-BIND	3.78e-133
6	771	48.4	125	1	HPAA_HELAC	NEURAMINYLLACTOSE-BIND	2.63e-118
7	107	6.7	616	1	ORC3_YEAST	ORIGIN RECOGNITION COM	4.42e-01
8	105	6.6	805	1	SYFB_MYCPN	PHENYLALANYL-TRNA SYNT	7.61e-01
9	105	6.6	1803	1	YJL3_YEAST	TRANSPOSON TY4 207.7 K	7.61e-01
10	104	6.5	313	1	MTNX_NEILA	CYTOSINE-SPECIFIC METH	9.95e-01
11	103	6.5	542	1	KPYK_YARLI	PYRUVATE KINASE (EC 2.	1.30e+00
12	104	6.5	616	1	Y329_METJA	HYPOTHETICAL PROTEIN M	9.95e-01
13	102	6.4	1142	1	KPC1_NEUCR	PROTEIN KINASE C-LIKE	1.69e+00
14	100	6.3	221	1	RL3_CHLTR	50S RIBOSOMAL PROTEIN	2.86e+00
15	101	6.3	767	1	YNY7_YEAST	PUTATIVE CYSTEINYL-TRN	2.20e+00
16	100	6.3	1251	1	RBP2_PLAVB	RETICULOCYTE BINDING P	2.86e+00
17	98	6.1	311	1	FMT_MYCPN	METHIONYL-TRNA FORMYLT	4.78e+00
18	98	6.1	330	1	YNJ7_YEAST	HYPOTHETICAL 37.0 KDA	4.78e+00
19	98	6.1	330	1	ETFA_CLOTS	ELECTRON TRANSFER FLAV	4.78e+00
20	98	6.1	705	1	YM37_YEAST	HYPOTHETICAL 81.5 KDA	4.78e+00
21	98	6.1	764	1	SYFB_HELPY	PHENYLALANYL-TRNA SYNT	4.78e+00
22	98	6.1	1858	1	P3K2_DICDI	PHOSPHATIDYLINOSITOL 3	4.78e+00
23	95	6.0	197	1	RUVA_BORBU	HOLLIDAY JUNCTION DNA	1.02e+01
24	95	6.0	296	1	COPE_YEAST	COATOMER EPSILON SUBUN	1.02e+01
25	96	6.0	368	1	K1CR_XENLA	KERATIN, TYPE I CYTOSK	7.94e+00
26	95	6.0	404	1	RFAL_SALTY	O-ANTIGEN LIGASE.	1.02e+01
27	95	6.0	436	1	YN95_YEAST	PUTATIVE 49.7 KDA MEMB	1.02e+01
28	96	6.0	671	1	FADH_ECOLI	2,4-DIENOYL-COA REDUCT	7.94e+00
29	95	6.0	845	1	SCP1_MESAU	SYNAPTONEMAL COMPLEX P	1.02e+01
30	96	6.0	866	1	LOX1_LENCU	LIPOXYGENASE (EC 1.13.	7.94e+00
31	95	6.0	917	1	YGJ3_YEAST	HYPOTHETICAL 104.8 KDA	1.02e+01
32	95	6.0	1140	1	XPE_CERAE	POSSIBLE DNA-REPAIR PR	1.02e+01
33	95	6.0	1570	1	P3K1_DICDI	PHOSPHATIDYLINOSITOL 3	1.02e+01
34	94	5.9	294	1	NUCG_MOUSE	ENDONUCLEASE G PRECURS	1.31e+01
35	94	5.9	338	1	RLA0_METTL	ACIDIC RIBOSOMAL PROTE	1.31e+01
36	94	5.9	476	1	PURA_WHEAT	ADENYLOSUCCINATE SYNTH	1.31e+01
37	94	5.9	501	1	TACY_PAEAL	ALVEOLYSIN PRECURSOR (	1.31e+01
38	94	5.9	1091	1	CIC2_HUMAN	DIHYDROPYRIDINE-SENSIT	1.31e+01
39	94	5.9	1181	1	YY02_METJA	HYPOTHETICAL PROTEIN M	1.31e+01
40	94	5.9	2415	1	SPCA_DROME	SPECTRIN ALPHA CHAIN.	1.31e+01
41	93	5.8	144	1	HAP3_YEAST	HAP3 TRANSCRIPTIONAL A	1.67e+01
42	93	5.8	186	1	PTH_HELPJ	PEPTIDYL-TRNA HYDROLAS	1.67e+01
43	93	5.8	402	1	PGK_CHLPN	PHOSPHOGLYCERATE KINAS	1.67e+01
44	93	5.8	842	1	HIS2_PICPA	HISTIDINE BIOSYNTHESIS	1.67e+01
45	93	5.8	3951	1	VGF1_IBVB	F1 PROTEIN.	1.67e+01

ALIGNMENTS

```
RESULT    1
ID   HPA3_HELPY     STANDARD;      PRT;   260 AA.
AC   Q48261;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N-
DE   ACETYLNEURAMINYLLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-
DE   BINDING SUBUNIT) (NLBH) (FLAGELLAR SHEATH ADHESIN).
GN   HPAA.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 11637;
RX   MEDLINE; 97431527.
RA   Jones A.C., Logan R.P., Foynes S., Cockayne A., Wren B.W., Penn C.W.;
RT   "A flagellar sheath protein of Helicobacter pylori is identical to
RT   HpaA, a putative N-acetylneuraminyllactose-binding hemagglutinin, but
RT   is not an adhesin for AGS cells.";
RL   J. Bacteriol. 179:5643-5647(1997).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC       ANCHOR (PROBABLE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X92502; CAA63246.1; -.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Flagella; Outer membrane; Lipoprotein; Signal.
FT   SIGNAL        1     27       BY SIMILARITY.
FT   CHAIN        28    260       NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ.
FT   LIPID        28     28       N-ACYL DIGLYCERIDE (PROBABLE).
FT   DOMAIN      134    139       N-ACETYL-NEURAMINYL-ALPHA(2,3)-LACTOSE
FT                                BINDING MOTIF (POTENTIAL).
SQ   SEQUENCE   260 AA;  29100 MW;  F613B93FF1FF1F40 CRC64;
```

Query Match 100.0%; Score 1594; DB 1; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.03e-279;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    2
ID   HPA0_HELPY     STANDARD;      PRT;   260 AA.
AC   P55969;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N-
DE   ACETYLNEURAMINYLLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-
DE   BINDING SUBUNIT) (NLBH) (FLAGELLAR SHEATH ADHESIN).
GN   HPAA OR HP0797.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE; 97394467.
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC       ANCHOR (PROBABLE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000591; AAD07844.1; -.
DR   TIGR; HP0797; -.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Flagella; Outer membrane; Lipoprotein; Signal.
FT   SIGNAL        1     27       BY SIMILARITY.
FT   CHAIN        28    260       NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ.
FT   LIPID        28     28       N-ACYL DIGLYCERIDE (PROBABLE).
SQ   SEQUENCE   260 AA;  29040 MW;  D9F86C8B0556D071 CRC64;

  Query Match           99.1%;  Score 1580;  DB 1;  Length 260;
  Best Local Similarity  98.7%;  Pred. No. 6.22e-277;
  Matches   230;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           ||||||||||||||||||||:|||:||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFVKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    3
ID   HPA2_HELPY     STANDARD;      PRT;   260 AA.
AC   Q48254;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N-
DE   ACETYLNEURAMINYLLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-
DE   BINDING SUBUNIT) (NLBH) (FLAGELLAR SHEATH ADHESIN).
GN   HPAA.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCUG 17874;
RX   MEDLINE; 96042077.
RA   O'Toole P.W., Janzon L., Doig P., Huang J., Kostrzynska M.,
RA   Trust T.J.;
RT   "The putative neuraminyllactose-binding hemagglutinin HpaA of
RT   Helicobacter pylori CCUG 17874 is a lipoprotein.";
RL   J. Bacteriol. 177:6049-6057(1995).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC       ANCHOR (PROBABLE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U35455; AAA79096.1; -.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Flagella; Outer membrane; Lipoprotein; Signal.
FT   SIGNAL        1     27       BY SIMILARITY.
FT   CHAIN        28    260       NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ.
FT   LIPID        28     28       N-ACYL DIGLYCERIDE (PROBABLE).
FT   DOMAIN      134    139       N-ACETYL-NEURAMINYL-ALPHA(2,3)-LACTOSE
FT                                BINDING MOTIF (POTENTIAL).
SQ   SEQUENCE   260 AA;  29003 MW;  676DDA952E99E03C CRC64;

  Query Match           99.0%;  Score 1578;  DB 1;  Length 260;
  Best Local Similarity  97.9%;  Pred. No. 1.55e-276;
  Matches   228;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTVLKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVINVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           ||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFVKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
```

```
Db    208 TMVKGTDNSNDAIKSALNKIFGSIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
          ||||||||||||||||||||::||||||||||||||||||||||||||||||||
Qy    208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    4
ID   HPA1_HELPY     STANDARD;      PRT;   260 AA.
AC   Q48264;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N-
DE   ACETYLNEURAMINYLLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-
DE   BINDING SUBUNIT) (NLBH) (FLAGELLAR SHEATH ADHESIN).
GN   HPAA.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=8826;
RX   MEDLINE; 93139035.
RA   Evans D.G., Karjalainen T.K., Evans D.J., Graham D.Y., Lee C.-H.;
RT   "Cloning, nucleotide sequence, and expression of a gene encoding an
RT   adhesin subunit protein of Helicobacter pylori.";
RL   J. Bacteriol. 175:674-683(1993).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC       ANCHOR (PROBABLE).
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X61574; CAA43773.1; -.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Flagella; Outer membrane; Lipoprotein; Signal.
FT   SIGNAL        1     27       BY SIMILARITY.
FT   CHAIN        28    260       NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ.
FT   LIPID        28     28       N-ACYL DIGLYCERIDE (PROBABLE).
FT   DOMAIN      134    139       N-ACETYL-NEURAMINYL-ALPHA(2,3)-LACTOSE
FT                                BINDING MOTIF (POTENTIAL).
SQ   SEQUENCE   260 AA;  29166 MW;  22489598065E7B14 CRC64;

  Query Match           97.3%;  Score 1551;  DB 1;  Length 260;
  Best Local Similarity  95.7%;  Pred. No. 3.53e-271;
  Matches   223;  Conservative    8;  Mismatches   2;  Indels   0;  Gaps   0;

Db     28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLKPAFQYSDNIAKEYENKFKNQTTLKV 87
          ||||||||||||||||||||||||||||||:||||||||||||||||||:|||
Qy     28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db     88 EEILQNQGYKVINVDSSDKDDFSFAQKKEGYLAVAMIGEIVLRPDPKRTIQKKSEPGLLF 147
          |:||||||||||:|||||||||||||||||||| ||||||||||||||||||||
Qy     88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db    148 STGLDKMEGVLIPAGFVKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
          ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Qy    148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db    208 TMVKGTDNSNDAIKSALNKIFASIMQEMDKKLTQRNLESYQKDAKELKNKRNR 260
          ||||||||||||||||||||:||||:||||||:|||||||||||| ||||
Qy    208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    5
ID   HPAA_HELNE     STANDARD;      PRT;   125 AA.
AC   Q48244;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ (N-ACETYLNEURAMINYLLACTOSE-
DE   BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-BINDING SUBUNIT) (NLBH)
DE   (FLAGELLAR SHEATH ADHESIN) (ADHESIN A) (FRAGMENT).
GN   HPAA OR HNAA.
OS   Helicobacter nemestrinae.
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96001251.
RA   Evans D.G., Lampert H.C., Nakano H., Eaton K.A., Burnens A.P.,
RA   Bronsdon M.A., Evans D.J. Jr.;
RT   "Genetic evidence for host specificity in the adhesin-encoding genes
RT   hxaA of Helicobacter acinonyx, hnaA of H. nemestrinae and hpaA of H.
RT   pylori.";
RL   Gene 163:97-102(1995).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC       ANCHOR (PROBABLE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U27127; AAA85564.1; -.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
KW   Flagella; Outer membrane; Lipoprotein.
FT   NON_TER       1      1
FT   DOMAIN       92     97       N-ACETYL-NEURAMINYL-ALPHA(2,3)-LACTOSE
FT                                BINDING MOTIF (POTENTIAL).
FT   NON_TER     125    125
SQ   SEQUENCE   125 AA;  14157 MW;  0EA29519F2B311AF CRC64;

  Query Match           53.2%;  Score 848;  DB 1;  Length 125;
  Best Local Similarity  96.0%;  Pred. No. 3.78e-133;
  Matches   120;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db      1 NYHPASEKVQALDEKILLLRPAFQYRDNIAKEYENKFKNQTTLKVEQILQNQGYKVINVD 60
          |||||||||||||||||||||||| |||||||||||||:|||||||||||||:||
Qy     43 NYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGYKVISVD 102

Db     61 SSDKDDLSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEGVLIPAG 120
          ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    103 SSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEGVLIPAG 162

Db    121 FVKVT 125
          |:|||
Qy    163 FIKVT 167


RESULT    6
ID   HPAA_HELAC     STANDARD;      PRT;   125 AA.
AC   Q47947;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ (N-ACETYLNEURAMINYLLACTOSE-
DE   BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-BINDING SUBUNIT) (NLBH)
DE   (FLAGELLAR SHEATH ADHESIN) (ADHESIN A) (FRAGMENT).
GN   HPAA OR HXAA.
OS   Helicobacter acinonyx.
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96001251.
RA   Evans D.G., Lampert H.C., Nakano H., Eaton K.A., Burnens A.P.,
RA   Bronsdon M.A., Evans D.J. Jr.;
```

```
RT   "Genetic evidence for host specificity in the adhesin-encoding genes
RT   hxaA of Helicobacter acinonyx, hnaA of H. nemestrinae and hpaA of H.
RT   pylori.";
RL   Gene 163:97-102(1995).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC       ANCHOR (PROBABLE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U27126; AAA85563.1; -.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
KW   Flagella; Outer membrane; Lipoprotein.
FT   NON_TER       1      1
FT   DOMAIN       92     97       N-ACETYL-NEURAMINYL-ALPHA(2,3)-LACTOSE
FT                                BINDING MOTIF (POTENTIAL).
FT   NON_TER     125    125
SQ   SEQUENCE   125 AA;  13923 MW;  8060097C2E6573E2 CRC64;

  Query Match           48.4%;  Score 771;  DB 1;  Length 125;
  Best Local Similarity  85.6%;  Pred. No. 2.63e-118;
  Matches  107;  Conservative   14;  Mismatches   4;  Indels   0;  Gaps   0;

Db       1 NYHPASETIQALDENILLLKPAFQYSDNVAKEYENKCKNQIALKVEEILQNQGYKVISVD 60
           ||||||| :|||||:||||:||||||||:||||||| ||| |||||:|||||||||||||
Qy      43 NYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGYKVISVD 102

Db      61 SSDKDDLSFAQKKEGYLTLSLSGEIVLRPDPKRTTQKKSEPGLLFSTGLDKMQGVLISAG 120
           ||||||:||||||||||::::::||||||||||| ||||||||||||||||:||||:||
Qy     103 SSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEGVLIPAG 162

Db     121 FVKVT 125
           |:|||
Qy     163 FIKVT 167


RESULT    7
ID   ORC3_YEAST     STANDARD;      PRT;   616 AA.
AC   P54790;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   ORIGIN RECOGNITION COMPLEX SUBUNIT 3 (ORIGIN RECOGNITION COMPLEX
DE   PROTEIN 62 KDA SUBUNIT).
GN   ORC3 OR OAF1 OR OIF1 OR YLL004W OR L1365.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE; 96069857.
RA   Bell S.P., Mitchell J., Leber J., Kobayashi R., Stillman B.;
RT   "The multidomain structure of Orc1p reveals similarity to regulators
RT   of DNA replication and transcriptional silencing.";
RL   Cell 83:563-568(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96239540.
RA   Hardy C.F.J.;
RT   "Characterization of an essential Orc2p-associated factor that plays
RT   a role in DNA replication.";
RL   Mol. Cell. Biol. 16:1832-1841(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY23;
RX   MEDLINE; 96405918.
RA   Miosga T., Zimmermann F.K.;
RT   "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on
RT   a 43.7 kb fragment of chromosome XII including an open reading frame
RT   homologous to the human cystic fibrosis transmembrane conductance
RT   regulator protein CFTR.";
RL   Yeast 12:693-708(1996).
CC   -!- FUNCTION: COMPONENT OF THE ORIGIN RECOGNITION COMPLEX (ORC) THAT
CC       BINDS ORIGINS OF REPLICATION. IT HAS A ROLE IN BOTH CHROMOSOMAL
CC       REPLICATION AND MATING TYPE TRANSCRIPTIONAL SILENCING. BINDS TO
CC       THE ARS CONSENSUS SEQUENCE (ACS) OF ORIGINS OF REPLICATION.
CC   -!- SUBUNIT: ORC IS COMPOSED OF SIX SUBUNITS OF 120 KDA, 71 KDA, 62
CC       KDA, 56 KDA, 53 KDA AND 50 KDA.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U34861; AAB38249.1; -.
DR   EMBL; U48888; AAC49270.1; -.
DR   EMBL; X91488; CAA62765.1; -.
DR   EMBL; Z73109; CAA97447.1; -.
DR   SGD; L0003074; ORC3.
KW   DNA replication; Nuclear protein; DNA-binding.
SQ   SEQUENCE   616 AA;  72077 MW;  3FAAEE3F645D3F9C CRC64;

  Query Match           6.7%;  Score 107;  DB 1;  Length 616;
  Best Local Similarity  25.7%;  Pred. No. 4.42e-01;
  Matches   26;  Conservative   33;  Mismatches  32;  Indels  10;  Gaps  10;

Db     197 KDLAMVFNFKDVDSINFNTLDNFIILLKSAFKYDHVKISLIFNINTNLSNIEKNLRQSTI 256
           :::|: :|:    | :  :||: |:|||::|| |   :|:  :: | :: | :  |:   |
Qy      36 NEVALKLNYHPA-SEKVQALDEKILLLRPAFQYSD-NIAKEYE-N-KFKN-QTALKVEQI 90

Db     257 RLLKRNYHKLDV-SSNKG-FKYGNQIFQSFLD-TVDGKLNL 294
             | :: |  : | ||:|  | ::    :::|  :::| : |
Qy      91 -LQNQGYKVISVDSSDKDDFSFAQKK-EGYLAVAMNGEIVL 129


RESULT    8
ID   SYFB_MYCPN     STANDARD;      PRT;   805 AA.
AC   P75563;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20) (PHENYLALANINE--
DE   TRNA LIGASE BETA CHAIN) (PHERS).
GN   PHET OR MP049.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE; 97105885.
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE + TRNA(PHE) = AMP +
CC       PYROPHOSPHATE + L-PHENYLALANYL-TRNA(PHE).
CC   -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000006; AAB95696.1; -.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
SQ   SEQUENCE   805 AA;  91713 MW;  937B7614E789547A CRC64;

  Query Match             6.6%;  Score 105;  DB 1;  Length 805;
  Best Local Similarity  28.4%;  Pred. No. 7.61e-01;
  Matches    23;  Conservative   21;  Mismatches  33;  Indels   4;  Gaps   4;

Db     506 ALQKLRTKLQTLGFHNVITYQLISPERARNFNLFGLSNLWEIKNP-LSNERSVLRVGLID 564
           :|:|:   |   || :|   : :| |   :| :  ||:| :| :  |   :|    ||:
Qy     150 GLDKMEGVLIPAGFIKVTILEPMSGESLDSFTM-DLSEL-DIQEKFLKTTHSSHSGGLVS 207

Db     565 SLLRVIQK-NAAYKNKLGNIF 584
           ::::  :: | | |: | :||
Qy     208 TMVKGTDNSNDAIKSALNKIF 228


RESULT    9
ID   YJL3_YEAST     STANDARD;      PRT;  1803 AA.
AC   P47024; P87192;
DT   01-FEB-1996 (Rel. 33, Created)
DT   15-FEB-2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   TRANSPOSON TY4 207.7 KDA HYPOTHETICAL PROTEIN.
GN   TY4B OR YJL113W OR J0780.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE; 97103775.
RA   Czieplucn C., Kordes E., Pujol A., Jauniaux J.-C.;
RT   "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
RT   reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,
RT   SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, CCT7, GZF3, two tRNA genes,
RT   three remnant delta elements and a Ty4 transposon.";
RL   Yeast 12:1471-1474(1996).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z49389; CAA89409.1; -.
KW   Transposable element; Hypothetical protein.
SQ   SEQUENCE   1803 AA;  207693 MW;  16DCD7284A8D52D3 CRC64;

  Query Match             6.6%;  Score 105;  DB 1;  Length 1803;
  Best Local Similarity  25.6%;  Pred. No. 7.61e-01;
  Matches    21;  Conservative   28;  Mismatches  29;  Indels   4;  Gaps   4;

Db    1171 NLKELKVTDKNVPTDNGTNVSPRLEQNIEASGSPVQTVNKSAFLNKEFSSLNMKRKRKRH 1230
           :| || : :| : | :::: :  :   : :: :: ::: ||| ||| |::: |    |:
Qy     183 DLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDNSNDAI-KSA-LNKIFANI-MQEIDKKL 239

Db    1231 DKNNSLTSYELERDKKRSKKNR 1252
             :| | ||: :    ::|:||
Qy     240 TQKN-LESYQKDAKELKGKRNR 260


RESULT   10
ID   MTNX_NEILA     STANDARD;      PRT;   313 AA.
AC   P24581;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CYTOSINE-SPECIFIC METHYLTRANSFERASE NLAX (EC 2.1.1.73) (M.NLAX).
OS   Neisseria lactamica.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23970 / NRCC 2118;
RX   MEDLINE; 91117164.
RA   Labbe D., Hoeltke H.J., Lau P.C.K.;
RT   "Cloning and characterization of two tandemly arranged DNA
RT   methyltransferase genes of Neisseria lactamica: an adenine-specific
RT   M.NlaIII and a cytosine-type methylase.";
RL   Mol. Gen. Genet. 224:101-110(1990).
CC   -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC       CCNGG, CAUSES SPECIFIC METHYLATION ON C-2 ON BOTH STRANDS.
CC   -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE =
CC       S-ADENOSYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.
CC   -!- SIMILARITY: STRONG, TO OTHER C5-DNA METHYLASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X54485; CAA38357.1; -.
DR   PIR; S12037; XYNHCL.
DR   HSSP; P05102; 5MHT.
DR   REBASE; RB02158; M.NlaX.
DR   PFAM; PF00145; DNA_methylase; 1.
DR   PRINTS; PR00105; C5METTRFRASE.
DR   PROSITE; PS00094; C5_MTASE_1; 1.
DR   PROSITE; PS00095; C5_MTASE_2; 1.
KW   Transferase; Methyltransferase; Restriction system.
FT   ACT_SITE     74      74       BY SIMILARITY.
SQ   SEQUENCE   313 AA;  34842 MW;  E75150C89ABEA611 CRC64;

  Query Match             6.5%;  Score 104;  DB 1;  Length 313;
  Best Local Similarity  34.9%;  Pred. No. 9.95e-01;
  Matches    29;  Conservative   18;  Mismatches  27;  Indels   9;  Gaps   9;

Db      98 ERILLAKKPQAFLL-ENV-KQLKGHDKGRT-LQVILAHLQQAGYKVYTEVLKARD-FGIP 153
           |:||| : | ||   :|: |:     | :| | |    ||  |||| :    :| |:::
Qy      56 EKILLLR-P-AFQYSDNIAKEYENKFKNQTALKVEQI-LQNQGYKVISVDSSDKDDFSFA 112

Db     154 QNRERIYLVGFLNHDVDFRFPQP 176
           |::|  ||   :| :: :| |:|
Qy     113 QKKEG-YLAVAMNGEIVLR-PDP 133


RESULT   11
ID   KPYK_YARLI     STANDARD;      PRT;   542 AA.
AC   P30614;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   PYRUVATE KINASE (EC 2.7.1.40).
OS   Yarrowia lipolytica (Candida lipolytica).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Dipodascaceae; Yarrowia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL Y-1094;
RX   MEDLINE; 92380493.
RA   Strick C.A., James L.C., O'Donnell M.M., Gollaher M.G., Franke A.E.;
RT   "The isolation and characterization of the pyruvate kinase-encoding
RT   gene from the yeast Yarrowia lipolytica.";
RL   Gene 118:65-72(1992).
RN   [2]
RP   ERRATUM.
RX   MEDLINE; 94171073.
RA   Strick C.A., James L.C., O'Donnell M.M., Gollaher M.G., Franke A.E.;
```

```
RL   Gene 140:141-143(1994).
CC   -!- CATALYTIC ACTIVITY: ATP + PYRUVATE = ADP + PHOSPHOENOLPYRUVATE.
CC   -!- COFACTOR: REQUIRES MAGNESIUM AND POTASSIUM.
CC   -!- PATHWAY: FINAL STEP IN GLYCOLYSIS.
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE PYRUVATE KINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M86863; AAA18520.1; -.
DR   PIR; JC1267; JC1267.
DR   HSSP; P00549; 1A3X.
DR   PFAM; PF00224; PK; 1.
DR   PRINTS; PR01050; PYRUVTKNASE.
DR   PROSITE; PS00110; PYRUVATE_KINASE; 1.
KW   Transferase; Kinase; Glycolysis; Magnesium.
FT   ACT_SITE     251     251       BY SIMILARITY.
FT   METAL        253     253       MAGNESIUM (POTENTIAL).
FT   METAL        274     274       MAGNESIUM (POTENTIAL).
FT   METAL        275     275       MAGNESIUM (POTENTIAL).
FT   BINDING      348     348       ADP (POTENTIAL).
SQ   SEQUENCE   542 AA;  59699 MW;  5BC44DDF0914BE77 CRC64;

  Query Match             6.5%;  Score 103;  DB 1;  Length 542;
  Best Local Similarity  23.1%;  Pred. No. 1.30e+00;
  Matches    27;  Conservative   31;  Mismatches  54;  Indels   5;  Gaps   5;

Db    157 DDGVLSFKVLEKIDGETLKVETLNNGKISSRKGVNLPGTDVDLPALSEKDKADLKFG-VE 215
          : |:|    |:|::|  :    ::   : :  | :| :  :||: |   :|  ||
Qy    142 EPGLLFSTGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKF-LKTTHSS 200

Db    216 HGVDMIFASFVR-TANDVQAIRDVLGEK-GKGIQVISKIENQQGVNNFDEILKETDG 270
          |:  :: :: |: | |  :||:  |    :: :| | |   |  :::::   ||  |
Qy    201 HSGGLV-STMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKG 256


RESULT   12
ID   Y329_METJA     STANDARD;      PRT;   616 AA.
AC   Q57775;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ0329.
GN   MJ0329.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE; 96337999.
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67487; AAB98317.1; -.
DR   TIGR; MJ0329; -.
KW   Hypothetical protein.
SQ   SEQUENCE   616 AA;  72036 MW;  2991EC36D4D64B82 CRC64;

  Query Match             6.5%;  Score 104;  DB 1;  Length 616;
  Best Local Similarity  25.7%;  Pred. No. 9.95e-01;
  Matches    19;  Conservative   26;  Mismatches  24;  Indels   5;  Gaps   5;

Db    363 GEILAAYP-DLTEKEV-EKLLKMAKEGK-GGLEYA-DEEKQKLLEKKRKSLEKIISK-LE 417
          || | ::  ||:| :: ||:|| :  :  |||  :     ::  :  : :|:||::: ::
Qy    174 GESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDNSNDAIKSALNKIFANIMQ 233

Db    418 KVGDKFGKKSMENF 431
          :  |:  |::|::
Qy    234 EIDKKLTQKNLESY 247


RESULT   13
ID   KPC1_NEUCR     STANDARD;      PRT;  1142 AA.
AC   P87253;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PROTEIN KINASE C-LIKE (EC 2.7.1.-).
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Arpaia G., Macino G.;
RT   "Molecular cloning of a protein kinase C homologue from Neurospora
RT   crassa.";
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC       BINDING DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       PKC SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y12002; CAA72731.1; -.
DR   PFAM; PF00130; DAG_PE-bind; 2.
DR   PFAM; PF00069; pkinase; 1.
DR   PFAM; PF00433; pkinase_C; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00479; DAG_PE_BINDING_DOMAIN; 2.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Zinc;
KW   Phorbol-ester binding; Duplication.
FT   DOMAIN       458     505       PHORBOL-ESTER AND DAG BINDING (BY
FT                                  SIMILARITY).
FT   DOMAIN       526     576       PHORBOL-ESTER AND DAG BINDING (BY
FT                                  SIMILARITY).
FT   DOMAIN       670     679       POLY-ALA.
FT   DOMAIN       714     717       POLY-GLN.
FT   DOMAIN       746     754       POLY-GLN.
FT   DOMAIN       817    1076       PROTEIN KINASE.
FT   NP_BIND      823     831       ATP (BY SIMILARITY).
FT   BINDING      846     846       ATP (BY SIMILARITY).
FT   ACT_SITE     942     942       BY SIMILARITY.
SQ   SEQUENCE  1142 AA;  127897 MW;  C599012D1D1F5970 CRC64;
```

```
  Query Match           6.4%;  Score 102;  DB 1;  Length 1142;
  Best Local Similarity  23.2%;  Pred. No. 1.69e+00;
  Matches    22;  Conservative   28;  Mismatches  39;  Indels   6;  Gaps   6;

Db    199 AAAKRVES-KQKIVLLKQALKRYEELHIDIDVDGPDDDSINLPALRKPLSGTLSIRILAV 257
          :|: :|::  :||:||: |: :| : :|  : :    :   |  :   |    : ::::|
Qy     46 PASEKVQALDEKILLLRPAF-QYSD-NIAKEYENKFKNQTALK-VEQILQNQ-GYKVISV 101

Db    258 KDVDHAPLGRFARSPETFIAVKAEDIVVARTKPSR 292
            |   :: ||:  | ::||  :  :| |  | |
Qy    102 DSSDKDDFS-FAQKKEGYLAVAMNGEIVLRPDPKR 135


RESULT   14
ID   RL3_CHLTR      STANDARD;      PRT;   221 AA.
AC   O84533;
DT   15-FEB-2000 (Rel. 39, Created)
DT   15-FEB-2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   50S RIBOSOMAL PROTEIN L3.
GN   RPLC OR RL3 OR CT528.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RX   MEDLINE; 99000809.
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
CC   -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND MAY
CC       PARTICIPATE IN THE FORMATION OF THE PEPTIDYLTRANSFERASE CENTER OF
CC       THE RIBOSOME (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE L3P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001323; AAC68129.1; -.
DR   PFAM; PF00297; Ribosomal_L3; 1.
DR   PROSITE; PS00474; RIBOSOMAL_L3; FALSE_NEG.
KW   Ribosomal protein; rRNA-binding.
SQ   SEQUENCE   221 AA;  23514 MW;  065BBC1462353F5B CRC64;

  Query Match           6.3%;  Score 100;  DB 1;  Length 221;
  Best Local Similarity  36.7%;  Pred. No. 2.86e+00;
  Matches    18;  Conservative   13;  Mismatches  15;  Indels   3;  Gaps   3;

Db     29 VISVDANVVAQLK-TASSDGYNAVQMGADVVQAPE-K-TIEKRFSKALL 74
          |||||::   ::  :   :|| || | :::|  |: | ||:|:   :||
Qy     98 VISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLL 146


RESULT   15
ID   YNY7_YEAST     STANDARD;      PRT;   767 AA.
AC   P53852;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PUTATIVE CYSTEINYL-TRNA SYNTHETASE C29E6.06C (EC 6.1.1.16) (CYSTEINE--
DE   TRNA LIGASE) (CYSRS).
GN   YNL247W OR N0885.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE; 97377992.
RA   Sen-Gupta M., Gueldener U., Beinhauer J.D., Fiedler T.A.,
RA   Hegemann J.H.;
RT   "Sequence analysis of the 33 kb long region between ORC5 and SUI1.
RT   from the left arm of chromosome XIV from Saccharomyces cerevisiae.";
RL   Yeast 13:849-860(1997).
RN   [2]
RP   SEQUENCE OF 44-767 FROM N.A.
RA   Poehlmann R., Philippsen P.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + L-CYSTEINE + TRNA(CYS) = AMP +
CC       PYROPHOSPHATE + L-CYSTEINYL-TRNA(CYS).
CC   -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X96722; CAA65497.1; -.
DR   EMBL; Z71523; CAA96154.1; -.
DR   PFAM; PF01406; tRNA-synt_1e; 1.
DR   PRINTS; PR00983; TRNASYNTHCYS.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; FALSE_NEG.
KW   Hypothetical protein; Aminoacyl-tRNA synthetase; Protein biosynthesis;
KW   Ligase; ATP-binding.
FT   SIMILAR      65     75       "HIGH" REGION.
FT   SIMILAR     427    431       "KMSKS" REGION.
FT   BINDING     430    430       ATP (BY SIMILARITY).
SQ   SEQUENCE   767 AA;  87530 MW;  495D526781B4BE2B CRC64;

  Query Match           6.3%;  Score 101;  DB 1;  Length 767;
  Best Local Similarity  27.7%;  Pred. No. 2.20e+00;
  Matches    18;  Conservative   16;  Mismatches  26;  Indels   5;  Gaps   4;

Db    101 QNVTDIDDKIILRARQNYLF-DNFVKENDTKFNATVVDKVKTALFQYINKNFTIQGSEIK 159
          ::| :|:||:|  |  : : ||: || : ||:     ||   | |  |  : : : :
Qy     49 EKVQALDEKILL-LRPAFQYSDNIAKEYENKFKNQTALKVEQIL-Q--NQGYKVISVDSS 104

Db    160 TIEEF 164
            ::|
Qy    105 DKDDF 109


Search completed: Tue Sep 26 16:16:16 2000
Job time : 13 secs.
```

****

****

.Psrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 26 16:18:57 2000; MasPar time 8.99 Seconds
613.770 Million cell updates/sec
Tabular output not generated.

Title: >US-09-308-435-4
Description: (28-260) from US09308435.pep
Perfect Score: 1599
Sequence: 1 CSPHIIETNEVALKLNYHPA.....QKNLESYQKDAKELKGKRNR 233

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 32.470; Variance 175.985; scale 0.185

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1599	100.0	260	1	W07450	Helicobacter pylori su	2.57e-113
2	1599	100.0	260	1	W60157	Helicobacter pylori 29	2.57e-113
3	1589	99.4	260	1	W60156	Helicobacter pylori 29	1.58e-112
4	1589	99.4	260	1	W07449	Helicobacter pylori su	1.58e-112
5	1589	99.4	461	1	W74466	Adhesin/V.cholerae tox	1.58e-112
6	1547	96.7	260	1	Y11000	H. pylori ORF 11ap2071	3.20e-109
7	1540	96.3	268	1	W20973	H. pylori derived prot	1.14e-108
8	302	18.9	97	1	W24663	H. pylori derived prot	5.74e-13
9	302	18.9	97	1	W20454	H. pylori derived prot	5.74e-13
10	235	14.7	249	1	W55445	H. pylori ORF 01ael101	3.94e-08
11	235	14.7	261	1	W55332	H. pylori ORF hp3el107	3.94e-08
12	227	14.2	249	1	W98332	H. pylori GHPO 1615 pr	1.46e-07
13	176	11.0	282	1	Y10951	H. pylori ORF 04eel110	5.19e-04
14	176	11.0	285	1	W20938	H. pylori secreted or	5.19e-04
15	118	7.4	147	1	W20468	H. pylori secreted or	3.41e+00
16	109	6.8	549	1	W21638	Grapevine leafroll vir	1.23e+01
17	107	6.7	615	1	R77276	ORC3 subunit of yeast	1.63e+01
18	107	6.7	615	1	W22226	S. cerevisiae origin o	1.63e+01
19	103	6.4	2039	1	W56322	Haemophilus paragallin	2.86e+01
20	103	6.4	2042	1	W56319	Haemophilus paragallin	2.86e+01
21	100	6.3	1254	1	R07503	Merozite apical-end-lo	4.33e+01
22	100	6.3	1254	1	W24575	Merozoite apical-end p	4.33e+01
23	99	6.2	1285	1	P93089	Pasteurella multocida	4.97e+01
24	97	6.1	119	1	W49813	Amino acid sequence of	6.53e+01
25	97	6.1	222	1	R39267	Humanised C4G1 Ig heav	6.53e+01
26	97	6.1	222	1	W49817	Fragment of humanised	6.53e+01
27	97	6.1	235	1	R39268	Humanised C4G1 Ig heav	6.53e+01
28	97	6.1	235	1	W49818	Amino acid sequence of	6.53e+01
29	97	6.1	449	1	R43339	Completely humanised C	6.53e+01
30	97	6.1	449	1	W49816	Amino acid sequence of	6.53e+01
31	97	6.1	552	1	W90172	Human heart muscle spe	6.53e+01
32	97	6.1	914	1	W24800	Spinocerebellar ataxia	6.53e+01
33	97	6.1	1312	1	W33807	Human ataxin-2.	6.53e+01
34	97	6.1	1313	1	W60213	Spinocerebellar ataxia	6.53e+01
35	96	6.0	246	1	W95018	Petunia MADS box trans	7.48e+01
36	96	6.0	947	1	R76708	Recombinant cold-resis	7.48e+01
37	96	6.0	947	1	R90924	Corn derived pyruvate	7.48e+01
38	94	5.9	175	1	R62783	Borrelia VS461 antigen	9.81e+01
39	95	5.9	599	1	W73479	Grapevine leafroll vir	8.57e+01
40	95	5.9	751	1	W13491	Helicobacter ClpB (N-t	8.57e+01
41	95	5.9	764	1	W98549	H. pylori GHPO 1604 pr	8.57e+01
42	94	5.9	1084	1	R71015	Human neuronal calcium	9.81e+01
43	94	5.9	1091	1	R71011	Human neuronal calcium	9.81e+01
44	95	5.9	1140	1	R72386	XAP-1, part of the DNA	8.57e+01
45	94	5.9	1392	1	Y06999	Restin protein sequenc	9.81e+01

ALIGNMENTS

RESULT 1
ID W07450 standard; Protein; 260 AA.
AC W07450;
DT 04-MAR-1997 (first entry)
DE Helicobacter pylori surface-exposed 29 kDa antigen.
KW Antigen; adhesin; immunogen; vaccine; immunisation; ulcer;
KW therapy; diagnosis.
OS Helicobacter pylori strain CCUG 17874 (NTCC 11637).
FH Key Location/Qualifiers
FT peptide 1. .27
FT /label= Sig_peptide
FT cleavage_site 25. .28
FT /note= "consensus recognition site for signal
FT peptidase II in prolipoproteins"
FT protein 28. .260
FT /label= Mat_protein
FT /note= "either the uncleaved or mature protein can
FT be used in applns. of the invention"
PN WO9638475-A1.
PD 05-DEC-1996.
PF 03-JUN-1996; SE0727.
PR 01-JUN-1995; SE-002007.
PR 21-MAR-1996; SE-001085.
PA (ASTR) ASTRA AB.
PI Bolin I, Svennerholm A;
DR WPI; 97-034307/03.
DR N-PSDB; T43835.
PT Helicobacter pylori surface exposed antigen - useful for treatment,
PT prophylaxis or diagnosis of mammalian infections
PS Claim 2; Page 37-38; 49pp; English.
CC A 29 kDa antigen (W07450) of Helicobacter pylori is an adhesin
CC expressed on the surface of resting and dividing forms of the
CC bacterium. It is conserved in all tested strains of H. pylori,
CC and gives rise to both systemic and local (mucosal) prodn. of
CC antibodies. Its amino acid sequence was deduced from a cDNA clone
CC (T43835); a sequence (W07549) deduced from a second cDNA clone
CC (T43834) differs only at position 222 (Ser for Arg). Recombinant
CC antigen can be produced in transformed host cells. It is useful
CC for the diagnosis and treatment of H. pylori infections and for
CC use as a vaccine.
SQ Sequence 260 AA;

Query Match 100.0%; Score 1599; DB 1; Length 260;
Best Local Similarity 100.0%; Pred. No. 2.57e-113;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

```
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260

Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
```

RESULT 2
ID W60157 standard; Protein; 260 AA.
AC W60157;
DT 28-SEP-1998 (first entry)
DE Helicobacter pylori 29 kDa surface-exposed antigen.
KW Cell surface antigen; adhesin; infection; vaccine; drug delivery.
OS Helicobacter pylori.
WO9822135-A1.
28-MAY-1998.
PF 18-NOV-1997; SE1927.
PR 22-NOV-1996; SE-004296.
PA (ASTR) ASTRA AB.
PI Berglindh T, Lofroth J;
DR WPI; 98-312174/27.
DR N-PSDB; V34643.
PT Aggregate of negatively charged lipid and Helicobacter pylori
PT antigen - useful in therapeutic or preventative vaccines against
PT both forms of bacterium, to generate systemic and mucosal antibody
PT responses
PS Disclosure; Page 35-36; 51pp; English.
CC This polypeptide comprises a 29 kDa surface antigen which is
CC expressed on the surface of both dividing (bacillary) and resting
CC (colloid) forms of Helicobacter pylori. The antigen gives rise to
CC both systemic and local (mucosal) production of antibodies. It is
CC an adhesin that is conserved in all tested strains of H. pylori,
CC and is also a putative virulence factor. The invention relates to
CC a new pharmaceutical composition comprising a lipid aggregate of a
CC negatively charged lipid or lipid mixture and at least one
CC antigenic, native or recombinant antigen of H. pylori, preferably
CC the surface-exposed 29 kDa antigen, or a nucleic acid (see V34643)
CC coding for such an antigen. The new formulations, which are useful
CC as vaccine formulations, elicit a protective immune response
CC against H. pylori infections, and are suitable for therapeutic and
CC prophylactic use. An adequate response is produced even without
CC adjuvant. A second, claimed 29 kDa antigen (see W60156) is
CC provided that differs only at residue 222 (Ser for Arg).
SQ Sequence 260 AA;

Query Match 100.0%; Score 1599; DB 1; Length 260;
Best Local Similarity 100.0%; Pred. No. 2.57e-113;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260

Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
```

RESULT 3
ID W60156 standard; Protein; 260 AA.
AC W60156;
DT 28-SEP-1998 (first entry)
DE Helicobacter pylori 29 kDa surface-exposed antigen.
KW Cell surface antigen; adhesin; infection; vaccine; drug delivery.
OS Helicobacter pylori.
PN WO9822135-A1.
PD 28-MAY-1998.
PF 18-NOV-1997; SE1927.
PR 22-NOV-1996; SE-004296.
PA (ASTR) ASTRA AB.
PI Berglindh T, Lofroth J;
DR WPI; 98-312174/27.
DR N-PSDB; V34642.
PT Aggregate of negatively charged lipid and Helicobacter pylori
PT antigen - useful in therapeutic or preventative vaccines against
PT both forms of bacterium, to generate systemic and mucosal antibody
PT responses
PS Claim 6; Page 28-30; 51pp; English.
CC This polypeptide comprises a 29 kDa surface antigen which is
CC expressed on the surface of both dividing (bacillary) and resting
CC (colloid) forms of Helicobacter pylori. The antigen gives rise to
CC both systemic and local (mucosal) production of antibodies. It is
CC an adhesin that is conserved in all tested strains of H. pylori,
CC and is also a putative virulence factor. The invention relates to
CC a new pharmaceutical composition comprising a lipid aggregate of a
CC negatively charged lipid or lipid mixture and at least one
CC antigenic, native or recombinant antigen of H. pylori, preferably
CC the surface-exposed 29 kDa antigen, or a nucleic acid (see V34642)
CC coding for such an antigen. The new formulations, which are useful
CC as vaccine formulations, elicit a protective immune response
CC against H. pylori infections, and are suitable for therapeutic and
CC prophylactic use. An adequate response is produced even without
CC adjuvant. A second 29 kDa antigen (see W60157) is provided that
CC differs only at residue 222 (Arg for Ser).
SQ Sequence 260 AA;

Query Match 99.4%; Score 1589; DB 1; Length 260;
Best Local Similarity 99.6%; Pred. No. 1.58e-112;
Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260

Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
```

RESULT 4
ID W07449 standard; Protein; 260 AA.
AC W07449;
DT 04-MAR-1997 (first entry)
DE Helicobacter pylori surface-exposed 29 kDa antigen.
KW Antigen; adhesin; immunogen; vaccine; immunisation; ulcer;
KW therapy; diagnosis.
OS Helicobacter pylori strain CCUG 17874 (NTCC 11637).
FH Key Location/Qualifiers
FT peptide 1. .27
FT /label= Sig_peptide
FT cleavage_site 25. .28

```
FT                   /note= "consensus recognition site for signal
FT                   peptidase II in prolipoproteins"
FT   protein         28. .260
FT                   /label= Mat_protein
FT                   /note= "either the uncleaved or mature protein can
FT                   be used in applns. of the invention"
PN   WO9638475-A1.
PD   05-DEC-1996.
PF   03-JUN-1996; SE0727.
PR   01-JUN-1995; SE-002007.
PR   21-MAR-1996; SE-001085.
PA   (ASTR ) ASTRA AB.
PI   Bolin I,  Svennerholm A;
DR   WPI; 97-034307/03.
DR   N-PSDB; T43834.
PT   Helicobacter pylori surface exposed antigen - useful for treatment,
PT   prophylaxis or diagnosis of mammalian infections
PS   Claim 2; Page 34-35; 49pp; English.
CC   A 29 kDa antigen (W07449) of Helicobacter pylori is an adhesin
CC   expressed on the surface of resting and dividing forms of the
CC   bacterium.  It is conserved in all tested strains of H. pylori,
CC   and gives rise to both systemic and local (mucosal) prodn. of
CC   antibodies.  Its amino acid sequence was deduced from a cDNA clone
CC   (T43834); a sequence (W07550) deduced from a second cDNA clone
CC   (T43835) differs only at position 222 (Arg for Ser).  Recombinant
CC   antigen can be produced in transformed host cells.  It is useful
CC   for the diagnosis and treatment of H. pylori infections and for
CC   use as a vaccine.
SQ   Sequence   260 AA;

  Query Match           99.4%;  Score 1589;  DB 1;  Length 260;
  Best Local Similarity  99.6%;  Pred. No. 1.58e-112;
  Matches   232;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    5
ID   W74466 standard; Protein; 461 AA.
AC   W74466;
DT   18-MAY-1999  (first entry)
DE   Adhesin/V.cholerae toxin A2 and B subunit fusion protein.
KW   Fusion gene; H. pylori infection; adhesin gene; V. cholerae toxin;
KW   A2 subunit; B subunit; diagnosis; vaccine; gastritis; gastric ulcer;
KW   duodenal ulcer.
OS   Chimeric - Helicobacter pylori.
OS   Chimeric - Vibrio cholerae.
PN   WO9853082-A1.
PD   26-NOV-1998.
PF   21-MAY-1997; 000091.
PR   21-MAY-1997; WO-KR0091.
PA   (DAEW-) DAEWOONG PHARM CO LTD.
PI   Kim BO, Lee BK, Park SK, Yoon SW, Yu YH;
DR   WPI; 98-434925/37.
DR   N-PSDB; X21955.
PT   Fusion gene consisting of the Helicobacter pylori adhesin gene
PT   ligated to the A2 and B subunit gene of Vibrio cholerae toxin -
PT   useful in the diagnosis and treatment of H. pylori
PS   Claim 2; Page 20-21; 32pp; English.
CC   This sequence represents the fusion protein of the invention, between the
CC   H. pylori adhesin gene, and the V. cholerae toxin A2 and B subunit genes.
CC   This protein is used in the diagnosis of and as a vaccine against
CC   H. pylori, which is the main cause of gastritis and gastric/duodenal
CC   ulcers. It may also be used in the production of an anti-H. pylori
CC   antibody. Vectors containing the DNA sequence and E. coli strains
CC   containing it form the basis of a process to produce the protein, which
CC   consists of culturing the microorganism and obtaining the protein. The
CC   protein is an effective vaccine against H. pylori due to its excellent
CC   immunogenicity for H. pylori, stability within the stomach environment,
CC   and its ability to penetrate through the stomach mucous membrane to
CC   stimulate sIgA production.
CC   Note: This sequence was indexed from WO9853082, which is the first
CC   major country equivalent to KR97059278.
SQ   Sequence   461 AA;

  Query Match           99.4%;  Score 1589;  DB 1;  Length 461;
  Best Local Similarity  99.6%;  Pred. No. 1.58e-112;
  Matches   232;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      39 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 98
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      99 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 158
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     159 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 218
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     219 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 271
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    6
ID   Y11000 standard; Protein; 260 AA.
AC   Y11000;
DT   08-JUN-1999  (first entry)
DE   H. pylori ORF 11ap20714_4797137_f3_45 cell envelope protein.
KW   Vaccine; probe; diagnostic; ORF; cell envelope protein;
KW   secreted protein; cellular protein.
OS   Helicobacter pylori.
PN   WO9818323-A1.
PD   07-MAY-1998.
PF   28-OCT-1997; U19575.
PR   14-JUL-1997; US-891928.
PR   28-OCT-1996; US-739150.
PR   06-DEC-1996; US-759739.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 98-271811/24.
DR   N-PSDB; X30467.
PT   Helicobacter pylori nucleic acids and proteins - used to develop
PT   products for the detection, prevention and treatment of H. pylori
PT   infections
PS   Claims 27, 31; Page 207-208; 279pp; English.
CC   Recombinant or substantially pure preparations of H. pylori polypeptides
CC   are disclosed, together with the nucleic acids encoding them. In all,
CC   73 ORFs are shown. The proteins are variously cell envelope proteins,
CC   secreted proteins or other cellular proteins. Vaccines containing the
CC   nucleic acids or proteins are claimed, as are probes containing at least
CC   8 nucleotides from the nucleic acid sequences. The vaccines are useful
CC   for treating or reducing the risk of H. pylori infections, and the
CC   probes can be used diagnostically for detecting the presence of
CC   Helicobacter in a sample. The products are also of use in screening
CC   for compounds having the ability to interfere with the H. pylori life
CC   cycle or to inhibit H. pylori infection.
SQ   Sequence   260 AA;

  Query Match           96.7%;  Score 1547;  DB 1;  Length 260;
```

```
  Best Local Similarity  95.7%;  Pred. No. 3.20e-109;
  Matches   223;  Conservative    7;  Mismatches   3;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTTLKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EEILQNQGYKVINVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           |:||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMERVLIPAGFVKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           |||||||| ||||||:|||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFASIMQEMDKKLTQRNLESYQKDAKELKNKRNR 260
           ||||||||||||| ||||||:||||:||||||:|||||||||||| ||||
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    7
ID   W20973 standard; Protein; 268 AA.
AC   W20973;
     21-JUL-1997  (first entry)
     H. pylori derived protein.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacterium; life cycle; activator;
KW   bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW   diagnosis.
OS   Helicobacter pylori.
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   N-PSDB; T68226.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Disclosure; Page 1363-64; 1481pp; English.
CC   This sequence represents a H. pylori protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
SQ   Sequence   268 AA;

  Query Match           96.3%;  Score 1540;  DB 1;  Length 268;
  Best Local Similarity  95.3%;  Pred. No. 1.14e-108;
  Matches   222;  Conservative    7;  Mismatches   4;  Indels   0;  Gaps   0;

Db      36 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTTLKV 95
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      96 EEILQNQGYKVINVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 155
           |:||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     156 STGLDKMERVLIPAGFVKVTILKPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 215
           |||||||| ||||||:||||| ||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     216 TMVKGTDNSNDAIKSALNKIFASIMQEMDKKLTQRNLESYQKDAKELKNKRNR 268
           |||||||||||||| |||||||:||||:||||||:|||||||||||| ||||
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    8
ID   W24663 standard; Protein; 97 AA.
AC   W24663;
DT   12-AUG-1997  (first entry)
DE   H. pylori derived protein.
KW   Transmembrane; cytoplasmic; cell envelope; flagella; transport;
KW   secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
KW   activator; inhibitor; bacterial life cycle; vaccine; immunise;
KW   detection; antisense; inhibition.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   misc_difference 94
FT                   /note= "encoded by GAK"
PN   WO9719098-A1.
PD   29-MAY-1997.
PF   15-NOV-1996; U18542.
PR   17-NOV-1995; US-561469.
PA   (ASTR ) ASTRA AB.
PI   Smith DH;
DR   WPI; 97-298052/27.
DR   N-PSDB; T77481.
PT   Helicobacter pylori nucleic acid sequences and related proteins -
PT   used for diagnostics and therapeutics
PS   Claim 18; Page 178; 235pp; English.
CC   This sequence represents an H. pylori derived protein, no further
CC   details are given in the specification.
CC   Helicobacter pylori has been strongly linked to chronic gastritis and
CC   duodenal ulcer disease.  The nucleic acid sequences of the invention
CC   are used to evaluate compounds, especially activators or inhibitors of
CC   bacterial life cycle, for the ability to bind an H. pylori nucleic acid
CC   sequence.  The nucleic acid sequences, and corresponding proteins, are
CC   also useful for generating vaccines for immunising subjects against H.
CC   pylori or for use in detecting the presence of Helicobacter species in
CC   a sample.  Antisense nucleic acid sequences of these sequences are
CC   used to inhibit expression of a gene from Helicobacter species.  H.
CC   pylori whole genomic DNA was isolated and nebulised to a median size of
CC   2000 bp.  Purified DNA fragments were blunt-ended and ligated to unique
CC   BstXI-linker adapters in 100-1000 fold molar excess.  These linkers are
CC   complementary to the BstXI-cut pMPX vectors, while the overhang is not
CC   self-complementary.  Therefore the linkers will not concatemerise nor
CC   will the cut vector re-ligate itself easily.  The linker-adapter inserts
CC   were ligated to each of the 20 pMPX vectors to construct a series of
CC   shotgun subclone libraries.  The purified DNA samples were then
CC   sequenced.
CC   Note: The ORF/protein reference number for this sequence was obtained
CC   from the related specification, WO9640893.
SQ   Sequence   97 AA;

  Query Match           18.9%;  Score 302;  DB 1;  Length 97;
  Best Local Similarity  97.7%;  Pred. No. 5.74e-13;
  Matches    43;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      53 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSXNI 96
           |||||||||||||||||||||||||||||||||||||||||||| ||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNI 71


RESULT    9
ID   W20454 standard; protein; 97 AA.
AC   W20454;
DT   29-JUL-1997  (first entry)
DE   H. pylori derived protein.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS   Helicobacter pylori.
```

```
FH   Key             Location/Qualifiers
FT   misc_difference 94
FT                   /label= Unknown
FT                   /note= "encoded by GAK"
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   N-PSDB; T67801.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Disclosure; Page 625; 1481pp; English.
CC   The present sequence is a H. pylori derived protein, no further details
CC   are given in the specification.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
SQ   Sequence   97 AA;

  Query Match           18.9%;  Score 302;  DB 1;  Length 97;
  Best Local Similarity  97.7%;  Pred. No. 5.74e-13;
  Matches    43;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      53 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSXNI 96
           |||||||||||||||||||||||||||||||||||||||||||| ||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNI 71


RESULT   10
W55445 standard; Protein; 249 AA.
W55445;
DT   24-JUN-1998  (first entry)
DE   H. pylori ORF 01ael1010_40688_c2_38 cell envelope OMP.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS   Helicobacter pylori.
PN   WO9737044-A1.
PD   09-OCT-1997.
PF   27-MAR-1997; U05223.
PR   06-DEC-1996; US-761318.
PR   29-MAR-1996; US-625811.
PR   02-APR-1996; US-758731.
PR   25-OCT-1996; US-736905.
PR   28-OCT-1996; US-738859.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 97-503122/46.
DR   N-PSDB; V24854.
PT   Helicobacter pylori nucleic acid sequences and encoded
PT   polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT   infection and for diagnosis of H. pylori infection
PS   Claims 14,80; Page 653; 1145pp; English.
CC   This sequence is a H. pylori cell envelope outer membrane protein (OMP).
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors. The
CC   DNA and probes derived from it may be used for the identification of
CC   H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC   acid sequences complementary to the DNA act as antisense sequences and
CC   can be used to prevent the translation of H. pylori mRNA. Antibodies
CC   against the protein can be used in immunoassays to evaluate the abundance
CC   and distribution of H. pylori-specific antigens. The genomic sequence of
CC   H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC   by mechanically shearing the bacterial DNA. The sequences were analysed
CC   for ORF of at least 180 nucleotides, and the predicted coding regions
CC   defined by computer evaluation. To identify likely H. pylori antigens for
CC   vaccine development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences of
CC   interest, particular regions can be isolated from H. pylori by PCR
CC   amplification for recombinant polypeptide production, e.g. in E. coli
CC   hosts.
SQ   Sequence   249 AA;

  Query Match           14.7%;  Score 235;  DB 1;  Length 249;
  Best Local Similarity  25.1%;  Pred. No. 3.94e-08;
  Matches    57;  Conservative   75;  Mismatches  73;  Indels  22;  Gaps  21;

Db      40 VELHFHY-PIKGKQEPKNNHLVVLIDPKIEANKVIPENYQKEFEKSLFLQLSNFLERKGY 98
           | | ::| |   | :: ::  ::|: | :: :  |: :|:: | :   | :  :|:  ||
Qy      38 VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db      99 SV-S-QFKDVSEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E-EKVID 147
           | | :  |  :::   |  : | : |:|:: :  :    ::  |::  | |   :|:
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     148 -M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTFVHRIK 203
            : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: | | :
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TMV-KGT 213

Db     204 ETDHDRAIKKIMNQAYHKVMAHITKELSKKHMERYEKVSSEMK-KRK 249
           :  :| |||: :|  : ::|  | | |: |::| |:| : |:| ||:
Qy     214 DNSND-AIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259


RESULT   11
ID   W55332 standard; Protein; 261 AA.
AC   W55332;
DT   15-JUN-1998  (first entry)
DE   H. pylori ORF hp3e11075orf3 protein.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   Misc_difference 261
FT                   /note= "the nucleotides encoding this amino acid
FT                         are not given in the specification"
PN   WO9737044-A1.
PD   09-OCT-1997.
PF   27-MAR-1997; U05223.
PR   06-DEC-1996; US-761318.
PR   29-MAR-1996; US-625811.
PR   02-APR-1996; US-758731.
PR   25-OCT-1996; US-736905.
PR   28-OCT-1996; US-738859.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 97-503122/46.
DR   N-PSDB; V24741.
PT   Helicobacter pylori nucleic acid sequences and encoded
PT   polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT   infection and for diagnosis of H. pylori infection
PS   Claim 14; Pages 558-559; 1145pp; English.
CC   This sequence is a H. pylori protein of unspecified function.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors. The
CC   DNA and probes derived from it may be used for the identification of
CC   H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
```

```
CC   acid sequences complementary to the DNA act as antisense sequences and
CC   can be used to prevent the translation of H. pylori mRNA. Antibodies
CC   against the protein can be used in immunoassays to evaluate the abundance
CC   and distribution of H. pylori-specific antigens. The genomic sequence of
CC   H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC   by mechanically shearing the bacterial DNA. The sequences were analysed
CC   for ORF of at least 180 nucleotides, and the predicted coding regions
CC   defined by computer evaluation. To identify likely H. pylori antigens for
CC   vaccine development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences of
CC   interest, particular regions can be isolated from H. pylori by PCR
CC   amplification for recombinant polypeptide production, e.g. in E. coli
CC   hosts.
SQ   Sequence   261 AA;

  Query Match           14.7%;  Score 235;  DB 1;  Length 261;
  Best Local Similarity  25.1%;  Pred. No. 3.94e-08;
  Matches    57;  Conservative   75;  Mismatches  73;  Indels  22;  Gaps  21;

Db      52 VELHFHY-PIKGKQEPKNNHLVVLIDPKIEANKVIPENYQKEFEKSLFLQLSNFLERKGY 110
           | | ::| |   | :: ::  ::|: | :: :  |: :|:: | :   | :  :|:  ||
Qy      38 VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db     111 SV-S-QFKDVSEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E-EKVID 159
           | | :  |  :::   |  : | : |:|:: :  :    ::  |::  | |   :|:
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     160 -M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTFVHRIK 215
            : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: | | :
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TMV-KGT 213

Db     216 ETDHDRAIKKIMNQAYHKVMAHITKELSKKHMERYEKVSSEMK-KRK 261
           :  :| |||: :|  : ::|  | | |: |::| |:| : |:| ||:
Qy     214 DNSND-AIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259


RESULT   12
ID   W98332 standard; Protein; 249 AA.
AC   W98332;
DT   31-MAR-1999  (first entry)
DE   H. pylori GHPO 1615 protein.
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease.
OS   Helicobacter pylori.
PN   WO9843478-A1.
PD   08-OCT-1998.
PF   01-APR-1998; U06371.
PR   29-JUL-1997; US-902615.
PR   01-APR-1997; US-833457.
PR   24-JUN-1997; US-881227.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI   Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR   WPI; 98-542293/46.
DR   N-PSDB; X14051.
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
PS   Claim 8; Page 455-456; 2054pp; English.
CC   This sequence represents a Helicobacter pylori GHPO protein of the
CC   invention. The polypeptides can be used for preventing or treating
CC   Helicobacter infections, and gastroduodenal diseases associated with
CC   these infections, including acute, chronic, and atrophic gastritis, and
CC   peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC   used for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
SQ   Sequence   249 AA;

  Query Match           14.2%;  Score 227;  DB 1;  Length 249;
  Best Local Similarity  25.0%;  Pred. No. 1.46e-07;
  Matches    58;  Conservative   75;  Mismatches  76;  Indels  23;  Gaps  22;

Db      35 NTGESVELHFHY-PIKGKQEPKNSHLVVLIEPKIEINKVIPESYQKEFEKSLFLQLSSFL 93
           :| | | | ::| |   | :: :   ::|: | :: :  |:  |:: | :   | :  :|
Qy      34 ETNE-VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQIL 91

Db      94 ERKGYSV-S-QFKDASEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E- 142
           :  || | | :  |  :::   |  : | :|:|:: :  :    ::  |::  | |
Qy      92 QNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGL 151

Db     143 EKVID-M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTF 198
           :|:   : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: |
Qy     152 DKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TM 209

Db     199 VHRIKETDHDQAIRKIMNQAYHKVMVHITKELSKKHMEHYEKVSSEMK-KRK 249
           | :  :  :| ||:: :|  : ::|  | | |: |::| |:| : |:| ||:
Qy     210 V-KGTDNSND-AIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259


RESULT   13
ID   Y10951 standard; Protein; 282 AA.
AC   Y10951;
DT   08-JUN-1999  (first entry)
DE   H. pylori ORF 04ee11108_3906963_f1_7 cell envelope protein.
KW   Vaccine; probe; diagnostic; ORF; cell envelope protein;
KW   secreted protein; cellular protein.
OS   Helicobacter pylori.
PN   WO9818323-A1.
PD   07-MAY-1998.
PF   28-OCT-1997; U19575.
PR   14-JUL-1997; US-891928.
PR   28-OCT-1996; US-739150.
PR   06-DEC-1996; US-759739.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 98-271811/24.
DR   N-PSDB; X30418.
PT   Helicobacter pylori nucleic acids and proteins - used to develop
PT   products for the detection, prevention and treatment of H. pylori
PT   infections
PS   Claims 27, 31; Page 151; 279pp; English.
CC   Recombinant or substantially pure preparations of H. pylori polypeptides
CC   are disclosed, together with the nucleic acids encoding them. In all,
CC   73 ORFs are shown. The proteins are variously cell envelope proteins,
CC   secreted proteins or other cellular proteins. Vaccines containing the
CC   nucleic acids or proteins are claimed, as are probes containing at least
CC   8 nucleotides from the nucleic acid sequences. The vaccines are useful
CC   for treating or reducing the risk of H. pylori infections, and the
CC   probes can be used diagnostically for detecting the presence of
CC   Helicobacter in a sample. The products are also of use in screening
CC   for compounds having the ability to interfere with the H. pylori life
CC   cycle or to inhibit H. pylori infection.
SQ   Sequence   282 AA;

  Query Match           11.0%;  Score 176;  DB 1;  Length 282;
  Best Local Similarity  25.1%;  Pred. No. 5.19e-04;
  Matches    56;  Conservative   59;  Mismatches  90;  Indels  18;  Gaps  14;

Db      64 PLDFNY-PV-HIVQAPQNHHVVGILMPRIQVSDNL-KPYIDKFQDALINQIQTIFEKRGY 120
           :| :|| |    |||  :  :: :| | :| |||: | | :||  :     :: |::::||
Qy      39 ALKLNYHPASEKVQA-LDEKIL-LLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db     121 QVLRFQD-EKA-LNVQDKKKIFSVLDLKGWVGILEDLKMNL--K-DPN---SPNLDTL-- 170
            |:  :  :|  ::  :||  :  : ::| : :  | |  :  | :|    | || :
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     171 VDQSSGSVWFNFYEPESNRVVHDFAVEVGTFQAITYTYTSTNNASGGFNSSKSVIHENLD 230
           |  ::| :   : || |   :  |::::: :: |   :  | ::| :     :::    |
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSELD-IQEKFLKTTHSSHSGGLVSTMVKGT-D 214

Db     231 KNREDAIHKILNRMYAVVMKKAVTELTKENIAKYRDAIDRMKG 273
           | :||| : ||:::| :|      ||  |:  |:     :||
Qy     215 -NSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKG 256
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 26 16:20:02 2000; MasPar time 9.81 Seconds
735.665 Million cell updates/sec
Tabular output not generated.

Title: >US-09-308-435-4
Description: (28-260) from US09308435.pep
Perfect Score: 1599
Sequence: 1 CSPHIIETNEVALKLNYHPA.....QKNLESYQKDAKELKGKRNR 233

Scoring table: PAM 150
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 48.644; Variance 103.561; scale 0.470

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1589	99.4	260	1	HPA3_HELPY	NEURAMINYLLACTOSE-BIND	5.31e-278
2	1575	98.5	260	1	HPA0_HELPY	NEURAMINYLLACTOSE-BIND	3.14e-275
3	1573	98.4	260	1	HPA2_HELPY	NEURAMINYLLACTOSE-BIND	7.82e-275
4	1546	96.7	260	1	HPA1_HELPY	NEURAMINYLLACTOSE-BIND	1.73e-269
5	848	53.0	125	1	HPAA_HELNE	NEURAMINYLLACTOSE-BIND	8.64e-133
6	771	48.2	125	1	HPAA_HELAC	NEURAMINYLLACTOSE-BIND	5.54e-118
7	109	6.8	616	1	Y329_METJA	HYPOTHETICAL PROTEIN M	2.68e-01
8	107	6.7	616	1	ORC3_YEAST	ORIGIN RECOGNITION COM	4.63e-01
9	104	6.5	313	1	MTNX_NEILA	CYTOSINE-SPECIFIC METH	1.04e+00
10	103	6.4	805	1	SYFB_MYCPN	PHENYLALANYL-TRNA SYNT	1.35e+00
11	102	6.4	1142	1	KPC1_NEUCR	PROTEIN KINASE C-LIKE	1.76e+00
12	100	6.3	221	1	RL3_CHLTR	50S RIBOSOMAL PROTEIN	2.97e+00
13	101	6.3	767	1	YNY7_YEAST	PUTATIVE CYSTEINYL-TRN	2.29e+00
14	100	6.3	1251	1	RBP2_PLAVB	RETICULOCYTE BINDING P	2.97e+00
15	100	6.3	1803	1	YJL3_YEAST	TRANSPOSON TY4 207.7 K	2.97e+00
16	99	6.2	542	1	KPYK_YARLI	PYRUVATE KINASE (EC 2.	3.85e+00
17	99	6.2	866	1	LOX1_LENCU	LIPOXYGENASE (EC 1.13.	3.85e+00
18	98	6.1	311	1	FMT_MYCPN	METHIONYL-TRNA FORMYLT	4.97e+00
19	97	6.1	365	1	RRN9_YEAST	RNA POLYMERASE I SPECI	6.40e+00
20	97	6.1	491	1	PBP_BACSU	PUTATIVE PENICILLIN BI	6.40e+00
21	97	6.1	730	1	EF2_METMT	ELONGATION FACTOR 2 (E	6.40e+00
22	97	6.1	773	1	MAK5_YEAST	ATP-DEPENDENT RNA HELI	6.40e+00
23	98	6.1	1858	1	P3K2_DICDI	PHOSPHATIDYLINOSITOL 3	4.97e+00
24	96	6.0	330	1	YNJ7_YEAST	HYPOTHETICAL 37.0 KDA	8.24e+00
25	96	6.0	330	1	ETFA_CLOTS	ELECTRON TRANSFER FLAV	8.24e+00
26	96	6.0	671	1	FADH_ECOLI	2,4-DIENOYL-COA REDUCT	8.24e+00
27	96	6.0	705	1	YM37_YEAST	HYPOTHETICAL 81.5 KDA	8.24e+00
28	96	6.0	845	1	SCP1_MESAU	SYNAPTONEMAL COMPLEX P	8.24e+00
29	96	6.0	947	1	PODK_MAIZE	PYRUVATE,PHOSPHATE DIK	8.24e+00
30	95	5.9	197	1	RUVA_BORBU	HOLLIDAY JUNCTION DNA	1.06e+01
31	94	5.9	294	1	NUCG_MOUSE	ENDONUCLEASE G PRECURS	1.35e+01
32	95	5.9	296	1	COPE_YEAST	COATOMER EPSILON SUBUN	1.06e+01
33	95	5.9	404	1	RFAL_SALTY	O-ANTIGEN LIGASE.	1.06e+01
34	95	5.9	436	1	YN95_YEAST	PUTATIVE 49.7 KDA MEMB	1.06e+01
35	94	5.9	476	1	PURA_WHEAT	ADENYLOSUCCINATE SYNTH	1.35e+01
36	94	5.9	501	1	TACY_PAEAL	ALVEOLYSIN PRECURSOR (	1.35e+01
37	95	5.9	764	1	SYFB_HELPY	PHENYLALANYL-TRNA SYNT	1.06e+01
38	95	5.9	917	1	YGJ3_YEAST	HYPOTHETICAL 104.8 KDA	1.06e+01
39	94	5.9	1091	1	CIC2_HUMAN	DIHYDROPYRIDINE-SENSIT	1.35e+01
40	95	5.9	1140	1	XPE_CERAE	POSSIBLE DNA-REPAIR PR	1.06e+01
41	94	5.9	1181	1	YY02_METJA	HYPOTHETICAL PROTEIN M	1.35e+01
42	94	5.9	1427	1	REST_HUMAN	RESTIN (CYTOPLASMIC LI	1.35e+01
43	95	5.9	1570	1	P3K1_DICDI	PHOSPHATIDYLINOSITOL 3	1.06e+01
44	94	5.9	1679	1	YIO9_YEAST	HYPOTHETICAL 195.1 KDA	1.35e+01
45	93	5.8	845	1	CC47_YEAST	CELL DIVISION CONTROL	1.73e+01

ALIGNMENTS

RESULT 1
ID HPA3_HELPY STANDARD; PRT; 260 AA.
AC Q48261;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N-
DE ACETYLNEURAMINYLLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-
DE BINDING SUBUNIT) (NLBH) (FLAGELLAR SHEATH ADHESIN).
GN HPAA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11637;
RX MEDLINE; 97431527.
RA Jones A.C., Logan R.P., Foynes S., Cockayne A., Wren B.W., Penn C.W.;
RT "A flagellar sheath protein of Helicobacter pylori is identical to
RT HpaA, a putative N-acetylneuraminyllactose-binding hemagglutinin, but
RT is not an adhesin for AGS cells.";
RL J. Bacteriol. 179:5643-5647(1997).
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC ANCHOR (PROBABLE).
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; X92502; CAA63246.1; -.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Flagella; Outer membrane; Lipoprotein; Signal.
FT SIGNAL 1 27 BY SIMILARITY.
FT CHAIN 28 260 NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ.
FT LIPID 28 28 N-ACYL DIGLYCERIDE (PROBABLE).
FT DOMAIN 134 139 N-ACETYL-NEURAMINYL-ALPHA(2,3)-LACTOSE
FT BINDING MOTIF (POTENTIAL).
SQ SEQUENCE 260 AA; 29100 MW; F613B93FF1FF1F40 CRC64;

Query Match 99.4%; Score 1589; DB 1; Length 260;
Best Local Similarity 99.6%; Pred. No. 5.31e-278;
Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
           ||||||||||||||| |||||||||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
```

```
RESULT    2
ID   HPA0_HELPY     STANDARD;      PRT;   260 AA.
AC   P55969;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N-
DE   ACETYLNEURAMINYLLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-
DE   BINDING SUBUNIT) (NLBH) (FLAGELLAR SHEATH ADHESIN).
GN   HPAA OR HP0797.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE; 97394467.
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC       ANCHOR (PROBABLE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000591; AAD07844.1; -.
DR   TIGR; HP0797; -.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Flagella; Outer membrane; Lipoprotein; Signal.
FT   SIGNAL        1     27       BY SIMILARITY.
FT   CHAIN        28    260       NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ.
FT   LIPID        28     28       N-ACYL DIGLYCERIDE (PROBABLE).
SQ   SEQUENCE   260 AA;  29040 MW;  D9F86C8B0556D071 CRC64;

  Query Match           98.5%;  Score 1575;  DB 1;  Length 260;
  Best Local Similarity  98.3%;  Pred. No. 3.14e-275;
  Matches   229;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87
```

```
Db      88 EQILQNQGYKVISVDSSDKDDLSFSQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           ||||||||||||||||||||:||:||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFVKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
           ||||||||||||||| |||||||||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
```

```
RESULT    3
ID   HPA2_HELPY     STANDARD;      PRT;   260 AA.
AC   Q48254;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N-
DE   ACETYLNEURAMINYLLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-
DE   BINDING SUBUNIT) (NLBH) (FLAGELLAR SHEATH ADHESIN).
GN   HPAA.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCUG 17874;
RX   MEDLINE; 96042077.
RA   O'Toole P.W., Janzon L., Doig P., Huang J., Kostrzynska M.,
RA   Trust T.J.;
RT   "The putative neuraminyllactose-binding hemagglutinin HpaA of
RT   Helicobacter pylori CCUG 17874 is a lipoprotein.";
RL   J. Bacteriol. 177:6049-6057(1995).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC       ANCHOR (PROBABLE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U35455; AAA79096.1; -.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Flagella; Outer membrane; Lipoprotein; Signal.
FT   SIGNAL        1     27       BY SIMILARITY.
FT   CHAIN        28    260       NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ.
FT   LIPID        28     28       N-ACYL DIGLYCERIDE (PROBABLE).
FT   DOMAIN      134    139       N-ACETYL-NEURAMINYL-ALPHA(2,3)-LACTOSE
FT                                BINDING MOTIF (POTENTIAL).
SQ   SEQUENCE   260 AA;  29003 MW;  676DDA952E99E03C CRC64;

  Query Match           98.4%;  Score 1573;  DB 1;  Length 260;
  Best Local Similarity  97.4%;  Pred. No. 7.82e-275;
  Matches   227;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTVLKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVINVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           ||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFVKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
```

```
Db     208 TMVKGTDNSNDAIKSALNKIFGSIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
           ||||||||||||||| ||||||::||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    4
ID   HPA1_HELPY     STANDARD;      PRT;   260 AA.
AC   Q48264;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N-
DE   ACETYLNEURAMINYLLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-
DE   BINDING SUBUNIT) (NLBH) (FLAGELLAR SHEATH ADHESIN).
GN   HPAA.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=8826;
RX   MEDLINE; 93139035.
RA   Evans D.G., Karjalainen T.K., Evans D.J., Graham D.Y., Lee C.-H.;
RT   "Cloning, nucleotide sequence, and expression of a gene encoding an
RT   adhesin subunit protein of Helicobacter pylori.";
RL   J. Bacteriol. 175:674-683(1993).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC       ANCHOR (PROBABLE).
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X61574; CAA43773.1; -.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Flagella; Outer membrane; Lipoprotein; Signal.
FT   SIGNAL        1     27       BY SIMILARITY.
FT   CHAIN        28    260       NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ.
FT   LIPID        28     28       N-ACYL DIGLYCERIDE (PROBABLE).
FT   DOMAIN      134    139       N-ACETYL-NEURAMINYL-ALPHA(2,3)-LACTOSE
FT                                BINDING MOTIF (POTENTIAL).
SQ   SEQUENCE   260 AA;  29166 MW;  22489598065E7B14 CRC64;

  Query Match           96.7%;  Score 1546;  DB 1;  Length 260;
  Best Local Similarity  95.3%;  Pred. No. 1.73e-269;
  Matches   222;  Conservative    8;  Mismatches   3;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLKPAFQYSDNIAKEYENKFKNQTTLKV 87
           ||||||||||||||||||||||||||||||||:||||||||||||||||||:|||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EEILQNQGYKVINVDSSDKDDFSFAQKKEGYLAVAMIGEIVLRPDPKRTIQKKSEPGLLF 147
           |:|||||||||:|||||||||||||||||||||| ||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFVKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFASIMQEMDKKLTQRNLESYQKDAKELKNKRNR 260
           |||||||||||||| |||||||:||||:||||||:|||||||||||||| ||||
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    5
ID   HPAA_HELNE     STANDARD;      PRT;   125 AA.
AC   Q48244;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ (N-ACETYLNEURAMINYLLACTOSE-
DE   BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-BINDING SUBUNIT) (NLBH)
DE   (FLAGELLAR SHEATH ADHESIN) (ADHESIN A) (FRAGMENT).
GN   HPAA OR HNAA.
OS   Helicobacter nemestrinae.
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96001251.
RA   Evans D.G., Lampert H.C., Nakano H., Eaton K.A., Burnens A.P.,
RA   Bronsdon M.A., Evans D.J. Jr.;
RT   "Genetic evidence for host specificity in the adhesin-encoding genes
RT   hxaA of Helicobacter acinonyx, hnaA of H. nemestrinae and hpaA of H.
RT   pylori.";
RL   Gene 163:97-102(1995).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC       ANCHOR (PROBABLE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U27127; AAA85564.1; -.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
KW   Flagella; Outer membrane; Lipoprotein.
FT   NON_TER       1      1
FT   DOMAIN       92     97       N-ACETYL-NEURAMINYL-ALPHA(2,3)-LACTOSE
FT                                BINDING MOTIF (POTENTIAL).
FT   NON_TER     125    125
SQ   SEQUENCE   125 AA;  14157 MW;  0EA29519F2B311AF CRC64;

  Query Match           53.0%;  Score 848;  DB 1;  Length 125;
  Best Local Similarity  96.0%;  Pred. No. 8.64e-133;
  Matches   120;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db       1 NYHPASEKVQALDEKILLLRPAFQYRDNIAKEYENKFKNQTTLKVEQILQNQGYKVINVD 60
           |||||||||||||||||||||||| ||||||||||||||:||||||||||||:||
Qy      43 NYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGYKVISVD 102

Db      61 SSDKDDLSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEGVLIPAG 120
           ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     103 SSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEGVLIPAG 162

Db     121 FVKVT 125
           |:|||
Qy     163 FIKVT 167


RESULT    6
ID   HPAA_HELAC     STANDARD;      PRT;   125 AA.
AC   Q47947;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ (N-ACETYLNEURAMINYLLACTOSE-
DE   BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-BINDING SUBUNIT) (NLBH)
DE   (FLAGELLAR SHEATH ADHESIN) (ADHESIN A) (FRAGMENT).
GN   HPAA OR HXAA.
OS   Helicobacter acinonyx.
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96001251.
RA   Evans D.G., Lampert H.C., Nakano H., Eaton K.A., Burnens A.P.,
RA   Bronsdon M.A., Evans D.J. Jr.;
```

```
RT   "Genetic evidence for host specificity in the adhesin-encoding genes
RT   hxaA of Helicobacter acinonyx, hnaA of H. nemestrinae and hpaA of H.
RT   pylori.";
RL   Gene 163:97-102(1995).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC       ANCHOR (PROBABLE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U27126; AAA85563.1; -.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
KW   Flagella; Outer membrane; Lipoprotein.
FT   NON_TER       1      1
FT   DOMAIN       92     97       N-ACETYL-NEURAMINYL-ALPHA(2,3)-LACTOSE
FT                                BINDING MOTIF (POTENTIAL).
FT   NON_TER     125    125
     SEQUENCE   125 AA;  13923 MW;  8060097C2E6573E2 CRC64;

  Query Match           48.2%;  Score 771;  DB 1;  Length 125;
  Best Local Similarity  85.6%;  Pred. No. 5.54e-118;
  Matches   107;  Conservative   14;  Mismatches   4;  Indels   0;  Gaps   0;

Db       1 NYHPASETIQALDENILLLKPAFQYSDNVAKEYENKCKNQIALKVEEILQNQGYKVISVD 60
           ||||||| :|||||:||||:||||||||:||||||| ||| |||||:||||||||||||||
Qy      43 NYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGYKVISVD 102

Db      61 SSDKDDLSFAQKKEGYLTLSLSGEIVLRPDPKRTTQKKSEPGLLFSTGLDKMQGVLISAG 120
           ||||||:|||||||||||:::::|||||||||||| |||||||||||||||||:||||:||
Qy     103 SSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEGVLIPAG 162

Db     121 FVKVT 125
           |:|||
Qy     163 FIKVT 167


RESULT    7
ID   Y329_METJA     STANDARD;      PRT;   616 AA.
AC   Q57775;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ0329.
GN   MJ0329.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE; 96337999.
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67487; AAB98317.1; -.
DR   TIGR; MJ0329; -.
KW   Hypothetical protein.
SQ   SEQUENCE   616 AA;  72036 MW;  2991EC36D4D64B82 CRC64;

  Query Match           6.8%;  Score 109;  DB 1;  Length 616;
  Best Local Similarity  25.7%;  Pred. No. 2.68e-01;
  Matches    19;  Conservative   27;  Mismatches  23;  Indels   5;  Gaps   5;

Db     363 GEILAAYP-DLTEKEV-EKLLKMAKEGK-GGLEYA-DEEKQKLLEKKRKSLEKIISK-LE 417
           || | ::  ||:| :: ||:|| :  :  |||  :     ::  :  :::|:||::: ::
Qy     174 GESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDNSNDAIKRALNKIFANIMQ 233

Db     418 KVGDKFGKKSMENF 431
             :  |:  |::|::
Qy     234 EIDKKLTQKNLESY 247


RESULT    8
ID   ORC3_YEAST     STANDARD;      PRT;   616 AA.
AC   P54790;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   ORIGIN RECOGNITION COMPLEX SUBUNIT 3 (ORIGIN RECOGNITION COMPLEX
DE   PROTEIN 62 KDA SUBUNIT).
GN   ORC3 OR OAF1 OR OIF1 OR YLL004W OR L1365.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE; 96069857.
RA   Bell S.P., Mitchell J., Leber J., Kobayashi R., Stillman B.;
RT   "The multidomain structure of Orc1p reveals similarity to regulators
RT   of DNA replication and transcriptional silencing.";
RL   Cell 83:563-568(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96239540.
RA   Hardy C.F.J.;
RT   "Characterization of an essential Orc2p-associated factor that plays
RT   a role in DNA replication.";
RL   Mol. Cell. Biol. 16:1832-1841(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY23;
RX   MEDLINE; 96405918.
RA   Miosga T., Zimmermann F.K.;
RT   "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on
RT   a 43.7 kb fragment of chromosome XII including an open reading frame
RT   homologous to the human cystic fibrosis transmembrane conductance
RT   regulator protein CFTR.";
RL   Yeast 12:693-708(1996).
CC   -!- FUNCTION: COMPONENT OF THE ORIGIN RECOGNITION COMPLEX (ORC) THAT
CC       BINDS ORIGINS OF REPLICATION. IT HAS A ROLE IN BOTH CHROMOSOMAL
CC       REPLICATION AND MATING TYPE TRANSCRIPTIONAL SILENCING. BINDS TO
CC       THE ARS CONSENSUS SEQUENCE (ACS) OF ORIGINS OF REPLICATION.
CC   -!- SUBUNIT: ORC IS COMPOSED OF SIX SUBUNITS OF 120 KDA, 71 KDA, 62
CC       KDA, 56 KDA, 53 KDA AND 50 KDA.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; U34861; AAB38249.1; -.
DR   EMBL; U48888; AAC49270.1; -.
DR   EMBL; X91488; CAA62765.1; -.
DR   EMBL; Z73109; CAA97447.1; -.
DR   SGD; L0003074; ORC3.
KW   DNA replication; Nuclear protein; DNA-binding.
SQ   SEQUENCE   616 AA;  72077 MW;  3FAAEE3F645D3F9C CRC64;

  Query Match             6.7%;  Score 107;  DB 1;  Length 616;
  Best Local Similarity  25.7%;  Pred. No. 4.63e-01;
  Matches    26;  Conservative   33;  Mismatches  32;  Indels  10;  Gaps  10;

Db    197 KDLAMVFNFKDVDSINFNTLDNFIILLKSAFKYDHVKISLIFNINTNLSNIEKNLRQSTI 256
          :::|: :|:    | :  :||: |:||::|| |   :|:  :: | :: | :  |:   |
Qy     36 NEVALKLNYHPA-SEKVQALDEKILLLRPAFQYSD-NIAKEYE-N-KFKN-QTALKVEQI 90

Db    257 RLLKRNYHKLDV-SSNKG-FKYGNQIFQSFLD-TVDGKLNL 294
          | :: |  : | ||:|  | ::    :::|  :::| : |
Qy     91 -LQNQGYKVISVDSSDKDDFSFAQKK-EGYLAVAMNGEIVL 129


RESULT    9
ID   MTNX_NEILA     STANDARD;      PRT;   313 AA.
AC   P24581;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CYTOSINE-SPECIFIC METHYLTRANSFERASE NLAX (EC 2.1.1.73) (M.NLAX).
OS   Neisseria lactamica.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23970 / NRCC 2118;
RX   MEDLINE; 91117164.
RA   Labbe D., Hoeltke H.J., Lau P.C.K.;
RT   "Cloning and characterization of two tandemly arranged DNA
RT   methyltransferase genes of Neisseria lactamica: an adenine-specific
RT   M.NlaIII and a cytosine-type methylase.";
RL   Mol. Gen. Genet. 224:101-110(1990).
CC   -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC       CCNGG, CAUSES SPECIFIC METHYLATION ON C-2 ON BOTH STRANDS.
CC   -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE =
CC       S-ADENOSYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.
CC   -!- SIMILARITY: STRONG, TO OTHER C5-DNA METHYLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54485; CAA38357.1; -.
DR   PIR; S12037; XYNHCL.
DR   HSSP; P05102; 5MHT.
DR   REBASE; RB02158; M.NlaX.
DR   PFAM; PF00145; DNA_methylase; 1.
DR   PRINTS; PR00105; C5METTRFRASE.
DR   PROSITE; PS00094; C5_MTASE_1; 1.
DR   PROSITE; PS00095; C5_MTASE_2; 1.
KW   Transferase; Methyltransferase; Restriction system.
FT   ACT_SITE     74      74       BY SIMILARITY.
SQ   SEQUENCE   313 AA;  34842 MW;  E75150C89ABEA611 CRC64;

  Query Match             6.5%;  Score 104;  DB 1;  Length 313;
  Best Local Similarity  34.9%;  Pred. No. 1.04e+00;
  Matches    29;  Conservative   18;  Mismatches  27;  Indels   9;  Gaps   9;

Db     98 ERILLAKKPQAFLL-ENV-KQLKGHDKGRT-LQVILAHLQQAGYKVYTEVLKARD-FGIP 153
          |:||| : | ||   :|: |:     | :| | |    ||  |||| :     :| |:::
Qy     56 EKILLLR-P-AFQYSDNIAKEYENKFKNQTALKVEQI-LQNQGYKVISVDSSDKDDFSFA 112

Db    154 QNRERIYLVGFLNHDVDFRFPQP 176
          |::|  ||   :| :: :| |:|
Qy    113 QKKEG-YLAVAMNGEIVLR-PDP 133


RESULT   10
ID   SYFB_MYCPN     STANDARD;      PRT;   805 AA.
AC   P75563;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20) (PHENYLALANINE--
DE   TRNA LIGASE BETA CHAIN) (PHERS).
GN   PHET OR MP049.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE; 97105885.
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE + TRNA(PHE) = AMP +
CC       PYROPHOSPHATE + L-PHENYLALANYL-TRNA(PHE).
CC   -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000006; AAB95696.1; -.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
SQ   SEQUENCE   805 AA;  91713 MW;  937B7614E789547A CRC64;

  Query Match             6.4%;  Score 103;  DB 1;  Length 805;
  Best Local Similarity  28.4%;  Pred. No. 1.35e+00;
  Matches    23;  Conservative   20;  Mismatches  34;  Indels   4;  Gaps   4;

Db    506 ALQKLRTKLQTLGFHNVITYQLISPERARNFNLFGLSNLWEIKNP-LSNERSVLRVGLID 564
          :|:|:   |   || :|   : :| | :| : ||:| :| : |  :|   ||:
Qy    150 GLDKMEGVLIPAGFIKVTILEPMSGESLDSFTM-DLSEL-DIQEKFLKTTHSSHSGGLVS 207

Db    565 SLLRVIQK-NAAYKNKLGNIF 584
          ::::  :: | | |  | :||
Qy    208 TMVKGTDNSNDAIKRALNKIF 228


RESULT   11
ID   KPC1_NEUCR     STANDARD;      PRT;  1142 AA.
AC   P87253;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PROTEIN KINASE C-LIKE (EC 2.7.1.-).
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Arpaia G., Macino G.;
RT   "Molecular cloning of a protein kinase C homologue from Neurospora
RT   crassa.";
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
```

```
CC       BINDING DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       PKC SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y12002; CAA72731.1; -.
DR   PFAM; PF00130; DAG_PE-bind; 2.
DR   PFAM; PF00069; pkinase; 1.
DR   PFAM; PF00433; pkinase_C; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00479; DAG_PE_BINDING_DOMAIN; 2.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Zinc;
KW   Phorbol-ester binding; Duplication.
FT   DOMAIN      458    505       PHORBOL-ESTER AND DAG BINDING (BY
FT                                SIMILARITY).
FT   DOMAIN      526    576       PHORBOL-ESTER AND DAG BINDING (BY
FT                                SIMILARITY).
FT   DOMAIN      670    679       POLY-ALA.
FT   DOMAIN      714    717       POLY-GLN.
FT   DOMAIN      746    754       POLY-GLN.
FT   DOMAIN      817   1076       PROTEIN KINASE.
FT   NP_BIND     823    831       ATP (BY SIMILARITY).
FT   BINDING     846    846       ATP (BY SIMILARITY).
FT   ACT_SITE    942    942       BY SIMILARITY.
SQ   SEQUENCE   1142 AA;  127897 MW;  C599012D1D1F5970 CRC64;

  Query Match             6.4%;  Score 102;  DB 1;  Length 1142;
  Best Local Similarity  23.2%;  Pred. No. 1.76e+00;
  Matches    22;  Conservative   28;  Mismatches  39;  Indels   6;  Gaps   6;

Db    199 AAAKRVES-KQKIVLLKQALKRYEELHIDIDVDGPDDDSINLPALRKPLSGTLSIRILAV 257
          :|: :|:: :|||:||: |: :| : :|  : :    :   |  :   |   : ::::|
Qy     46 PASEKVQALDEKILLLRPAF-QYSD-NIAKEYENKFKNQTALK-VEQILQNQ-GYKVISV 101

Db    258 KDVDHAPLGRFARSPETFIAVKAEDIVVARTKPSR 292
             |   :: ||:  | ::||  :  :| |  | |
Qy    102 DSSDKDDFS-FAQKKEGYLAVAMNGEIVLRPDPKR 135


RESULT   12
ID   RL3_CHLTR      STANDARD;      PRT;   221 AA.
AC   O84533;
DT   15-FEB-2000 (Rel. 39, Created)
DT   15-FEB-2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   50S RIBOSOMAL PROTEIN L3.
GN   RPLC OR RL3 OR CT528.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RX   MEDLINE; 99000809.
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
CC   -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND MAY
CC       PARTICIPATE IN THE FORMATION OF THE PEPTIDYLTRANSFERASE CENTER OF
CC       THE RIBOSOME (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE L3P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001323; AAC68129.1; -.
DR   PFAM; PF00297; Ribosomal_L3; 1.
DR   PROSITE; PS00474; RIBOSOMAL_L3; FALSE_NEG.
KW   Ribosomal protein; rRNA-binding.
SQ   SEQUENCE   221 AA;  23514 MW;  065BBC1462353F5B CRC64;

  Query Match             6.3%;  Score 100;  DB 1;  Length 221;
  Best Local Similarity  36.7%;  Pred. No. 2.97e+00;
  Matches    18;  Conservative   13;  Mismatches  15;  Indels   3;  Gaps   3;

Db     29 VISVDANVVAQLK-TASSDGYNAVQMGADVVQAPE-K-TIEKRFSKALL 74
          |||||::   ::  :   :|| || | :::|  |: | ||:|:   :||
Qy     98 VISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLL 146


RESULT   13
ID   YNY7_YEAST      STANDARD;      PRT;   767 AA.
AC   P53852;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PUTATIVE CYSTEINYL-TRNA SYNTHETASE C29E6.06C (EC 6.1.1.16) (CYSTEINE--
DE   TRNA LIGASE) (CYSRS).
GN   YNL247W OR N0885.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE; 97377992.
RA   Sen-Gupta M., Gueldener U., Beinhauer J.D., Fiedler T.A.,
RA   Hegemann J.H.;
RT   "Sequence analysis of the 33 kb long region between ORC5 and SUI1
RT   from the left arm of chromosome XIV from Saccharomyces cerevisiae.";
RL   Yeast 13:849-860(1997).
RN   [2]
RP   SEQUENCE OF 44-767 FROM N.A.
RA   Poehlmann R., Philippsen P.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + L-CYSTEINE + TRNA(CYS) = AMP +
CC       PYROPHOSPHATE + L-CYSTEINYL-TRNA(CYS).
CC   -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X96722; CAA65497.1; -.
DR   EMBL; Z71523; CAA96154.1; -.
DR   PFAM; PF01406; tRNA-synt_1e; 1.
DR   PRINTS; PR00983; TRNASYNTHCYS.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; FALSE_NEG.
KW   Hypothetical protein; Aminoacyl-tRNA synthetase; Protein biosynthesis;
KW   Ligase; ATP-binding.
FT   SIMILAR      65     75       "HIGH" REGION.
FT   SIMILAR     427    431       "KMSKS" REGION.
FT   BINDING     430    430       ATP (BY SIMILARITY).
SQ   SEQUENCE   767 AA;  87530 MW;  495D526781B4BE2B CRC64;

  Query Match             6.3%;  Score 101;  DB 1;  Length 767;
```

```
  Best Local Similarity  27.7%;  Pred. No. 2.29e+00;
  Matches    18;  Conservative   16;  Mismatches  26;  Indels   5;  Gaps   4;

Db    101 QNVTDIDDKIILRARQNYLF-DNFVKENDTKFNATVVDKVKTALFQYINKNFTIQGSEIK 159
          ::|  :|:||:|  |  : : ||: || : ||:     ||   | |  |  : : : :
Qy     49 EKVQALDEKILL-LRPAFQYSDNIAKEYENKFKNQTALKVEQIL-Q--NQGYKVISVDSS 104

Db    160 TIEEF 164
          ::|
Qy    105 DKDDF 109


RESULT   14
ID   RBP2_PLAVB     STANDARD;      PRT;  1251 AA.
AC   Q00799;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   RETICULOCYTE BINDING PROTEIN 2 (FRAGMENT).
GN   RBP2.
OS   Plasmodium vivax (strain Belem).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92315338.
RA   Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT   "A reticulocyte-binding protein complex of Plasmodium vivax
RT   merozoites.";
RL   Cell 69:1213-1226(1992).
CC   -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC       HUMAN RETICULOCYTE CELLS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND (PROBABLE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M88098; AAA29744.1; -.
KW   Malaria; Receptor; Membrane.
FT   NON_TER       1      1
 T   NON_TER    1251   1251
SQ   SEQUENCE   1251 AA;  143741 MW;  54BA51C7404AC572 CRC64;

  Query Match             6.3%;  Score 100;  DB 1;  Length 1251;
  Best Local Similarity  29.9%;  Pred. No. 2.97e+00;
  Matches    20;  Conservative   17;  Mismatches  26;  Indels   4;  Gaps   4;

Db    759 FESKNNVVLETENMSKNTNELDVHKNIQD-AYKV-ALEILAHSDEIDTKQKDSSKL-IEM 815
          |:  :|:  | ||: ||   | |   :|: :||| ::: :  |::   ||  : | : |
Qy     65 FQYSDNIAKEYENKFKNQTALKVEQILQNQGYKVISVDS-SDKDDFSFAQKKEGYLAVAM 123

Db    816 GNQIYLK 822
          :| |:
Qy    124 NGEIVLR 130


RESULT   15
ID   YJL3_YEAST     STANDARD;      PRT;  1803 AA.
AC   P47024; P87192;
DT   01-FEB-1996 (Rel. 33, Created)
DT   15-FEB-2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   TRANSPOSON TY4 207.7 KDA HYPOTHETICAL PROTEIN.
GN   TY4B OR YJL113W OR J0780.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE; 97103775.
RA   Cziepluch C., Kordes E., Pujol A., Jauniaux J.-C.;
RT   "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
RT   reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,
RT   SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, CCT7, GZF3, two tRNA genes,
RT   three remnant delta elements and a Ty4 transposon.";
RL   Yeast 12:1471-1474(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49389; CAA89409.1; -.
KW   Transposable element; Hypothetical protein.
SQ   SEQUENCE   1803 AA;  207693 MW;  16DCD7284A8D52D3 CRC64;

  Query Match             6.3%;  Score 100;  DB 1;  Length 1803;
  Best Local Similarity  24.4%;  Pred. No. 2.97e+00;
  Matches    20;  Conservative   28;  Mismatches  30;  Indels   4;  Gaps   4;

Db   1171 NLKELKVTDKNVPTDNGTNVSPRLEQNIEASGSPVQTVNKSAFLNKEFSSLNMKRKRKRH 1230
          :| || : :| : | :::: :  :   : :: :: ::: | | ||| |::: |    |:
Qy    183 DLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDNSNDAI-KRA-LNKIFANI-MQEIDKKL 239

Db   1231 DKNNSLTSYELERDKKRSKKNR 1252
            :| | ||: :    ::|:||
Qy    240 TQKN-LESYQKDAKELKGKRNR 260


Search completed: Tue Sep 26 16:20:13 2000
Job time : 11 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 26 16:15:28 2000; MasPar time 15.30 Seconds
718.193 Million cell updates/sec

Tabular output not generated.

Title: >US-09-308-435-2
Description: (28-260) from US09308435.pep
Perfect Score: 1594
Sequence: 1 CSPHIIETNEVALKLNYHPA.....QKNLESYQKDAKELKGKRNR 233

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir64
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 47.512; Variance 117.669; scale 0.404

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1580	99.1	260	2	E64619	adhesin A - Helicobac	1.16e-237
2	1564	98.1	260	2	C71896	neuraminyllactose-bin	6.15e-235
3	843	52.9	124	2	PC4089	adhesin A - Helicobac	2.25e-113
4	766	48.1	124	2	PC4090	adhesin A - Helicobac	1.26e-100
5	230	14.4	249	2	E71865	probable neuraminylla	4.54e-16
6	222	13.9	249	2	B64571	probable neuraminylla	6.09e-15
7	171	10.7	282	2	A71930	probable neuraminylla	5.00e-08
8	161	10.1	278	2	D64581	hypothetical protein	9.69e-07
9	110	6.9	1802	2	S52611	TyB protein - yeast (	1.00e+00
10	108	6.8	320	2	F75063	hypothetical protein	1.63e+00
11	107	6.7	616	2	S64746	ORC3 protein - yeast	2.07e+00
12	106	6.6	394	2	E70135	flagellar protein (fl	2.63e+00
13	105	6.6	805	2	S73374	phenylalanine--tRNA l	3.33e+00
14	105	6.6	1465	2	S31262	TyB protein - yeast (	3.33e+00
15	105	6.6	1803	2	S56894	TyB protein - yeast (	3.33e+00
16	104	6.5	313	1	XYNHCL	site-specific DNA-met	4.22e+00
17	104	6.5	417	2	T15174	hypothetical protein	4.22e+00
18	103	6.5	527	2	JC1267	pyruvate kinase (EC 2	5.34e+00
19	104	6.5	616	2	A64341	hypothetical protein	4.22e+00
20	102	6.4	559	2	T08523	transposase - Enterob	6.74e+00
21	102	6.4	559	2	S32177	transposase - Klebsie	6.74e+00
22	100	6.3	221	2	F71507	probable L3 ribosomal	1.07e+01
23	101	6.3	319	2	E70091	hypothetical protein	8.50e+00
24	101	6.3	471	2	S54460	hypothetical protein	8.50e+00
25	101	6.3	571	2	F70040	sulfite reductase hom	8.50e+00
26	101	6.3	592	2	G69159	sensory transduction	8.50e+00
27	101	6.3	767	2	S63220	probable membrane pro	8.50e+00
28	100	6.3	856	2	C71956	probable endopeptidas	1.07e+01
29	101	6.3	1000	2	T13636	probable minor struct	8.50e+00
30	100	6.3	1252	2	B42771	reticulocyte-binding	1.07e+01
31	99	6.2	405	2	T10260	patatin-like protein	1.34e+01
32	99	6.2	1134	2	D75014	hypothetical protein	1.34e+01
33	99	6.2	1467	2	PC1253	TyB protein - yeast (	1.34e+01
34	98	6.1	119	2	PL0089	Ig heavy chain V regi	1.69e+01
35	97	6.1	308	2	C75125	gmp synthase, cter do	2.11e+01
36	98	6.1	311	2	S73625	methionyl-tRNA formyl	1.69e+01
37	98	6.1	330	2	S58255	hypothetical protein	1.69e+01
38	98	6.1	705	2	S54521	probable membrane pro	1.69e+01
39	98	6.1	764	2	B64570	phenylalanine--tRNA l	1.69e+01
40	98	6.1	980	2	S54986	regulatory protein -	1.69e+01
41	97	6.1	9376	2	T14593	syringomycin syntheta	2.11e+01
42	96	6.0	231	2	D64486	hypothetical protein	2.64e+01
43	96	6.0	368	2	A28825	keratin, type I nonep	2.64e+01
44	96	6.0	672	2	F65096	probable NADH-depende	2.64e+01
45	96	6.0	1265	2	F71429	hypothetical protein	2.64e+01

ALIGNMENTS

```
RESULT    1
ENTRY             E64619     #type complete
TITLE             adhesin A - Helicobacter pylori (strains 26695 and 8826)
ALTERNATE_NAMES   flagellar sheath adhesin hpaA;
                    N-acetylneuraminyllactose-binding fibrillar hemagglutinin
                    receptor-binding subunit
ORGANISM          #formal_name Helicobacter pylori
DATE              09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
                    26-Feb-1999
ACCESSIONS        E64619; B47052; C47052; PC4088
REFERENCE         A64520
   #authors       Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
                    Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
                    H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
                    J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
                    Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
                    McKenney, K.; Fitzegerald, L.M.; Lee, N.; Adams, M.D.;
                    Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
                    Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
                    Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
                    W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
                    C.M.; Venter, J.C.
   #journal       Nature (1997) 388:539-547
   #title         The complete genome sequence of the gastric pathogen
                    Helicobacter pylori.
   #cross-references MUID:97394467
   #accession     E64619
      ##status         nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues       1-260 ##label TOM
      ##cross-references GB:AE000591; GB:AE000511; NID:g2313918; PID:g2313925;
                          TIGR:HP0797
      ##experimental_source strain 26695
REFERENCE         A47052
   #authors       Evans, D.G.; Karjalainen, T.K.; Evans Jr., D.J.; Graham,
                    D.Y.; Lee, C.H.
   #journal       J. Bacteriol. (1993) 175:674-683
   #title         Cloning, nucleotide sequence, and expression of a gene
                    encoding an adhesin subunit protein of Helicobacter pylori.
   #cross-references MUID:93139035
   #accession     B47052
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-2,'T',4,'G',6-18,'T',20-61,'K',63-83,'T',85-88,'E',
                        90-99,'N',101-108,'F',110-111,'A',113-152,'L',154-179,
                        'LEWI' ##label EVA1
      ##cross-references GB:X61574; NID:g732735
```

```
      ##note          sequence extracted from NCBI backbone (NCBIN:123532,
                        NCBIP:123534); sequence inconsistent with the
                        nucleotide translation
   #accession     C47052
      ##status        preliminary
      ##molecule_type DNA
      ##residues      'MR',204-218,'R',220,'LC',223-228,'RS',231-234,'M',
                        236-241,'R',243-255,'N',257-260 ##label EVA2
      ##cross-references GB:X61574; NID:g732735
      ##note          sequence extracted from NCBI backbone (NCBIN:123532,
                        NCBIP:123535)
REFERENCE         PC4088
   #authors       Evans, D.G.; Lampert, H.C.; Nakano, H.; Eaton, K.A.; Burnens,
                    A.P.; Bronsdon, M.A.; Evans Jr., D.J.
   #journal       Gene (1995) 163:97-102
   #title         Genetic evidence for host specificity in the adhesin-encoding
                    genes hxaA of Helicobacter acinonyx, hnaA of H. nemestrinae
                    and hpaA of H. pylori.
   #cross-references MUID:96001251
   #accession     PC4088
      ##molecule_type DNA
      ##residues      43-61,'K',63-83,'T',85-88,'E',90-99,'N',101-108,'F',
                        110-111,'A',113-123,'I',125-166 ##label EVA3
      ##experimental_source strain 8826
GENETICS
   #gene          hpaA; HP0797
FEATURE
   134-139             #region receptor binding #status predicted
SUMMARY           #length 260  #molecular-weight 29040  #checksum 2565

  Query Match           99.1%;  Score 1580;  DB 2;  Length 260;
  Best Local Similarity  98.7%;  Pred. No. 1.16e-237;
  Matches   230;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDLSFSQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           ||||||||||||||||||||:||:||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFVKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    2
ENTRY             C71896     #type complete
TITLE             neuraminyllactose-binding hemagglutinin precursor -
                    Helicobacter pylori (strain J99)
ORGANISM          #formal_name Helicobacter pylori
   #variety       strain J99
DATE              12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
                    12-Feb-1999
ACCESSIONS        C71896
REFERENCE         A71800
   #authors       Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
                    Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
                    B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
                    Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
                    Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
                    G.F.; Trust, T.J.
   #journal       Nature (1999) 397:176-180
   #title         Genomic sequence comparison of two unrelated isolates of the
                    human gastric pathogen Helicobacter pylori.
   #cross-references MUID:99120557
   #accession     C71896
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-260 ##label ARN
      ##cross-references GB:AE001504; GB:AE001439; NID:g4155285; PID:g4155289
      ##experimental_source strain J99
GENETICS
   #gene          hpaA
SUMMARY           #length 260  #molecular-weight 29199  #checksum 1435

  Query Match           98.1%;  Score 1564;  DB 2;  Length 260;
  Best Local Similarity  96.6%;  Pred. No. 6.15e-235;
  Matches   225;  Conservative    7;  Mismatches   1;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTTLKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EEILQNQGYKVINVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           |:|||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFVKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFASIMQEMDKKLTQRNLESYQKDAKELKNKRNR 260
           ||||||||||||||||||||:||||:||||||:||||||||||||| ||||
Qy     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    3
ENTRY             PC4089     #type fragment
TITLE             adhesin A - Helicobacter nemestrinae (fragment)
ALTERNATE_NAMES   HnaA protein
ORGANISM          #formal_name Helicobacter nemestrinae
DATE              14-Nov-1995 #sequence_revision 08-Feb-1996 #text_change
                    18-Sep-1998
ACCESSIONS        PC4089
REFERENCE         PC4088
   #authors       Evans, D.G.; Lampert, H.C.; Nakano, H.; Eaton, K.A.; Burnens,
                    A.P.; Bronsdon, M.A.; Evans Jr., D.J.
   #journal       Gene (1995) 163:97-102
   #title         Genetic evidence for host specificity in the adhesin-encoding
                    genes hxaA of Helicobacter acinonyx, hnaA of H. nemestrinae
                    and hpaA of H. pylori.
   #cross-references MUID:96001251
   #accession     PC4089
      ##molecule_type DNA
      ##residues      1-124 ##label EVA
      ##cross-references GB:U27127
GENETICS
   #gene          hnaA
FEATURE
   92-97               #region receptor binding #status predicted
SUMMARY           #length 124  #checksum 6484

  Query Match           52.9%;  Score 843;  DB 2;  Length 124;
  Best Local Similarity  96.0%;  Pred. No. 2.25e-113;
  Matches   119;  Conservative    4;  Mismatches   1;  Indels   0;  Gaps   0;

Db       1 NYHPASEKVQALDEKILLLRPAFQYRDNIAKEYENKFKNQTTLKVEQILQNQGYKVINVD 60
           |||||||||||||||||||||| ||||||||||||||:|||||||||||||||:||
Qy      43 NYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGYKVISVD 102

Db      61 SSDKDDLSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEGVLIPAG 120
           ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     103 SSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEGVLIPAG 162

Db     121 FVKV 124
           |:||
Qy     163 FIKV 166


RESULT    4
```

```
ENTRY            PC4090     #type fragment
TITLE            adhesin A - Helicobacter acinonyx (fragment)
ALTERNATE_NAMES  HxaA protein
ORGANISM         #formal_name Helicobacter acinonyx
DATE             14-Nov-1995 #sequence_revision 08-Feb-1996 #text_change
                   18-Sep-1998
ACCESSIONS       PC4090
REFERENCE        PC4088
   #authors      Evans, D.G.; Lampert, H.C.; Nakano, H.; Eaton, K.A.; Burnens,
                   A.P.; Bronsdon, M.A.; Evans Jr., D.J.
   #journal      Gene (1995) 163:97-102
   #title        Genetic evidence for host specificity in the adhesin-encoding
                   genes hxaA of Helicobacter acinonyx, hnaA of H. nemestrinae
                   and hpaA of H. pylori.
   #cross-references MUID:96001251
   #accession    PC4090
      ##molecule_type DNA
      ##residues      1-124 ##label EVA
      ##cross-references GB:U27126
GENETICS
   #gene         hxaA
FEATURE
   92-97              #region receptor binding #status predicted
SUMMARY          #length 124  #checksum 6624

  Query Match           48.1%;  Score 766;  DB 2;  Length 124;
  Best Local Similarity  85.5%;  Pred. No. 1.26e-100;
  Matches   106;  Conservative   14;  Mismatches   4;  Indels   0;  Gaps   0;

Db       1 NYHPASETIQALDENILLLKPAFQYSDNVAKEYENKCKNQIALKVEEILQNQGYKVISVD 60
           ||||||| :|||||:||||:||||||||:||||||| ||| |||||:|||||||||||||
Qy      43 NYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGYKVISVD 102

Db      61 SSDKDDLSFAQKKEGYLTLSLSGEIVLRPDPKRTTQKKSEPGLLFSTGLDKMQGVLISAG 120
           ||||||:||||||||||:::::|||||||||||| ||||||||||||||||:||||:||
Qy     103 SSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEGVLIPAG 162

Db     121 FVKV 124
           |:||
Qy     163 FIKV 166


RESULT    5
ENTRY            E71865     #type complete
 ITLE            probable neuraminyllactose-binding hemagglutinin jhp0971 -
                   Helicobacter pylori (strain J99)
ORGANISM         #formal_name Helicobacter pylori
   #variety      strain J99
DATE             12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
                   26-Feb-1999
ACCESSIONS       E71865
REFERENCE        A71800
   #authors      Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
                   Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
                   B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
                   Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
                   Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
                   G.F.; Trust, T.J.
   #journal      Nature (1999) 397:176-180
   #title        Genomic sequence comparison of two unrelated isolates of the
                   human gastric pathogen Helicobacter pylori.
   #cross-references MUID:99120557
   #accession    E71865
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-249 ##label ARN
      ##cross-references GB:AE001526; GB:AE001439; NID:g4155550; PID:g4155554
      ##experimental_source strain J99
GENETICS
   #gene         jhp0971
SUMMARY          #length 249  #molecular-weight 28513  #checksum 9499

  Query Match           14.4%;  Score 230;  DB 2;  Length 249;
  Best Local Similarity  25.1%;  Pred. No. 4.54e-16;
  Matches    57;  Conservative   74;  Mismatches  74;  Indels  22;  Gaps  21;

Db      40 VELHFHY-PIKGKQEPKNNHLVVLIDPKIEANKVIPENYQKEFEKSLFLQLSNFLERKGY 98
           | | ::| |   | :: ::  ::|: | :: :  |: :|:: | :   | : :|:  ||
Qy      38 VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db      99 SV-S-QFKDVSEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E-EKVID 147
           | | :  |  :::   |  : | : |:|:: :  :   ::  |::  | |   :|:
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     148 -M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTFVHRIK 203
            : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: | | :
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TMV-KGT 213

Db     204 ETDHDRAIKKIMNQAYHKVMAHITKELSKKHMERYEKVSSEMK-KRK 249
           :  :| |||  :|  : ::|  | | |: |::| |:| : |:| ||:
Qy     214 DNSND-AIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259


RESULT    6
ENTRY            B64571     #type complete
TITLE            probable neuraminyllactose-binding hemagglutinin HP0410 -
                   Helicobacter pylori (strain 26695)
ORGANISM         #formal_name Helicobacter pylori
DATE             09-Aug-1997 #sequence_revision 15-Aug-1997 #text_change
                   26-Feb-1999
ACCESSIONS       B64571
REFERENCE        A64520
   #authors      Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
                   Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
                   H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
                   J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
                   Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
                   McKenney, K.; Fitzegerald, L.M.; Lee, N.; Adams, M.D.;
                   Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
                   Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
                   Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
                   W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
                   C.M.; Venter, J.C.
   #journal      Nature (1997) 388:539-547
   #title        The complete genome sequence of the gastric pathogen
                   Helicobacter pylori.
   #cross-references MUID:97394467
   #accession    B64571
      ##status        preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues      1-249 ##label TOM
      ##cross-references GB:AE000557; GB:AE000511; NID:g2313514; PID:g2313516;
                        TIGR:HP0410
SUMMARY          #length 249  #molecular-weight 28349  #checksum 9957

  Query Match           13.9%;  Score 222;  DB 2;  Length 249;
  Best Local Similarity  25.0%;  Pred. No. 6.09e-15;
  Matches    58;  Conservative   74;  Mismatches  77;  Indels  23;  Gaps  22;

Db      35 NTGESVELHFHY-PIKGKQEPKNSHLVVLIEPKIEINKVIPESYQKEFEKSLFLQLSSFL 93
           :| | | | ::| |   | :: :   ::|: | :: :  |:  |:: | :   | :  :|
Qy      34 ETNE-VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQIL 91

Db      94 ERKGYSV-S-QFKDASEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E- 142
           :  || | | :  |  :::   |  : | : |:|:: :  :   ::  |::  | |
Qy      92 QNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGL 151

Db     143 EKVID-M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTF 198
           :|:   : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: |
Qy     152 DKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TM 209

Db     199 VHRIKETDHDQAIRKIMNQAYHKVMVHITKELSKKHMEHYEKVSSEMK-KRK 249
           | :  :  :| ||:  :|  : ::|  | | |: |::| |:| : |:| ||:
Qy     210 V-KGTDNSND-AIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259
```

```
RESULT    7
ENTRY             A71930     #type complete
TITLE             probable neuraminyllactose-binding hemagglutinin jhp0444 -
                    Helicobacter pylori (strain J99)
ORGANISM          #formal_name Helicobacter pylori
   #variety       strain J99
DATE              12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
                    26-Feb-1999
ACCESSIONS        A71930
REFERENCE         A71800
   #authors       Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
                    Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
                    B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
                    Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
                    Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
                    G.F.; Trust, T.J.
   #journal       Nature (1999) 397:176-180
   #title         Genomic sequence comparison of two unrelated isolates of the
                    human gastric pathogen Helicobacter pylori.
   #cross-references MUID:99120557
   #accession     A71930
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-282 ##label ARN
      ##cross-references GB:AE001479; GB:AE001439; NID:g4154979; PID:g4154991
      ##experimental_source strain J99
GENETICS
   #gene          jhp0444
SUMMARY           #length 282  #molecular-weight 32075  #checksum 2233

  Query Match           10.7%;  Score 171;  DB 2;  Length 282;
  Best Local Similarity  25.1%;  Pred. No. 5.00e-08;
  Matches    56;  Conservative   58;  Mismatches  91;  Indels  18;  Gaps  14;

Db      64 PLDFNY-PV-HIVQAPQNHHVVGILMPRIQVSDNL-KPYIDKFQDALINQIQTIFEKRGY 120
           :| :|| |    |||  :  :: :| | :| |||: | | :|| :     :: |::::||
Qy      39 ALKLNYHPASEKVQA-LDEKIL-LLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db     121 QVLRFQD-EKA-LNVQDKKKIFSVLDLKGWVGILEDLKMNL--K-DPN---SPNLDTL-- 170
            |:  :  :|  ::  :|:  :  : ::| : :  | |  :  | :|    |  || :
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     171 VDQSSGSVWFNFYEPESNRVVHDFAVEVGTFQAITYTYTSTNNASGGFNSSKSVIHENLD 230
           |  ::| :   : || |   :  |::::: :: |   :  | ::| :     :::    |
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSELD-IQEKFLKTTHSSHSGGLVSTMVKGT-D 214

Db     231 KNREDAIHKILNRMYAVVMKKAVTELTKENIAKYRDAIDRMKG 273
            | :|||   ||:::| :|      ||  |:  |:     :||
Qy     215 -NSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKG 256


 ESULT    8
ENTRY             D64581     #type complete
TITLE             hypothetical protein HP0492 - Helicobacter pylori (strain
                    26695)
ORGANISM          #formal_name Helicobacter pylori
DATE              09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
                    10-Oct-1997
ACCESSIONS        D64581
REFERENCE         A64520
   #authors       Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
                    Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
                    H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
                    J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
                    Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
                    McKenney, K.; Fitzegerald, L.M.; Lee, N.; Adams, M.D.;
                    Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
                    Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
                    Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
                    W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
                    C.M.; Venter, J.C.
   #journal       Nature (1997) 388:539-547
   #title         The complete genome sequence of the gastric pathogen
                    Helicobacter pylori.
   #cross-references MUID:97394467
   #accession     D64581
      ##status         preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues       1-278 ##label TOM
      ##cross-references GB:AE000564; GB:AE000511; NID:g2313602; PID:g2313613;
                        TIGR:HP0492
SUMMARY           #length 278  #molecular-weight 31947  #checksum 4354

  Query Match           10.1%;  Score 161;  DB 2;  Length 278;
  Best Local Similarity  25.0%;  Pred. No. 9.69e-07;
  Matches    56;  Conservative   61;  Mismatches  85;  Indels  22;  Gaps  18;

Db      62 PLDFNY-PI-HIVQAPQNHHVVGILTPRIQVSDNL-KPYIDKFQDALINQIQTIFEKRGY 118
           :| :|| |    |||  :  :: :| | :| |||: | | :|| :     :: |::::||
Qy      39 ALKLNYHPASEKVQA-LDEKIL-LLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db     119 QVLRFQD-EKA-LNAQDKRKIFSVLDLKGWVGILEDLKMNL--K-DPN---NPNLDTL-- 168
            |:  :  :|  ::  :|:  :  : ::| : :  | |  :  | :|    :  || :
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     169 VDQSSGSVWFNFYEPESNRVVHDFAVEVGTFQAM-TY-TYKHN-NSGGLNSSNSIIHEYL 225
           |  ::| :   : || |   :  |::::: ::    :    |: :|||| | :::
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGL-VS-TMVKG-T 213

Db     226 EKNKEDAIHKILNRMYAVVMKKAVTELTKENIDKYREAIDRMKG 269
           : | :|||   ||:::| :|      ||  |:: |:     :||
Qy     214 D-NSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKG 256


RESULT    9
ENTRY             S52611     #type complete
TITLE             TyB protein - yeast (Saccharomyces cerevisiae)
                    retrotransposon Ty4
ALTERNATE_NAMES   protein YHL008w-a
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              05-May-1995 #sequence_revision 19-Oct-1995 #text_change
                    08-Nov-1996
ACCESSIONS        S52611
REFERENCE         S46794
   #authors       Favello, T.
   #submission    submitted to the EMBL Data Library, June 1994
   #description   The sequence of S. cerevisiae cosmid L5018.
   #accession     S52611
      ##molecule_type DNA
      ##residues       1-1802 ##label FAV
      ##cross-references EMBL:U11581
GENETICS
   #map_position 8L
   #mobile_element retrotransposon Ty4
SUMMARY           #length 1802  #molecular-weight 207968  #checksum 3295

  Query Match            6.9%;  Score 110;  DB 2;  Length 1802;
  Best Local Similarity  26.8%;  Pred. No. 1.00e+00;
  Matches    22;  Conservative   27;  Mismatches  29;  Indels   4;  Gaps   4;

Db    1170 NLKELKVTDKNVPTDNGTNVSPRLEQNIEASGSPVQTVNKSAFLNKEFSSLNMKRKRKRH 1229
           :| || : :| : | :::: :  :   : :: :: ::: ||| ||| |::: |    |:
Qy     183 DLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDNSNDAI-KSA-LNKIFANI-MQEIDKKL 239

Db    1230 DKNNSLTSYELERDKKRSKRNR 1251
             :| | ||: :    ::||||
Qy     240 TQKN-LESYQKDAKELKGKRNR 260


RESULT   10
ENTRY             F75063     #type complete
TITLE             hypothetical protein PAB0993 - Pyrococcus abyssi (strain
                    Orsay)
ORGANISM          #formal_name Pyrococcus abyssi
```

```
DATE            20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
                  20-Aug-1999
ACCESSIONS      F75063
REFERENCE       A75001
   #authors     anonymous, Genoscope
   #submission  submitted to the EMBL Data Library, July 1999
   #description Pyrococcus abyssi genome sequence: insights into archaeal
                  chromosome structure and evolution.
   #accession   F75063
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-320 ##label KAW
      ##cross-references GB:AJ248287; GB:AL096836; NID:g5458657;
                        PIDN:CAB50403.1; PID:e1516301; PID:g5458916
      ##experimental_source strain Orsay
GENETICS
   #gene        PAB0993
SUMMARY         #length 320  #molecular-weight 37677  #checksum 8652

  Query Match             6.8%;  Score 108;  DB 2;  Length 320;
  Best Local Similarity  24.7%;  Pred. No. 1.63e+00;
  Matches    21;  Conservative   32;  Mismatches  25;  Indels   7;  Gaps   7;

Db     131 ERMFSFPLEIAE-SLEKRLIKSLEK-RVEELVEEKVKNSNNGISPEVIREFIDK-YDSLV 187
           | : || ::::|  :: :::|:    :   ||  :|| ::|: : : |:  ::| : :::
Qy     175 ESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDNS-N-DAIKSALNKIFANIM 232

Db     188 RE-NVEL-RRRLESREKIIKDLREK 210
           :| :  | :: ||| :|  |:|: |
Qy     233 QEIDKKLTQKNLESYQKDAKELKGK 257


RESULT   11
ENTRY           S64746     #type complete
TITLE           ORC3 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein L1365; protein YLL004w
ORGANISM        #formal_name Saccharomyces cerevisiae
DATE            01-Aug-1995 #sequence_revision 24-May-1996 #text_change
                  06-Feb-1998
ACCESSIONS      S64746; S70572
REFERENCE       S64743
   #authors     Miosga, T.; Zimmermann, F.K.
   #submission  submitted to the Protein Sequence Database, May 1996
   #accession   S64746
      ##molecule_type DNA
      ##residues      1-616 ##label MIO
      ##cross-references EMBL:Z73109; NID:g1360162; PID:e245444; PID:g1360163;
                        MIPS:YLL004w
      ##experimental_source strain S288C
REFERENCE       S70557
   #authors     Miosga, T.; Zimmermann, F.K.
   #journal     Yeast (1996) 12:693-708
   #title       Sequence analysis of the CEN12 region of Saccharomyces
                  cerevisiae on a 43.7 kb fragment of chromosome XII
                  including an open reading frame homologous to the human
                  cystic fibrosis transmembrane conductance regulator protein
                  CFTR.
   #cross-references MUID:96405918
   #accession   S70572
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-616 ##label MIW
      ##cross-references EMBL:X91488; NID:g1495203; PID:e199016; PID:g1495220
      ##note          the nucleotide sequence was submitted to the EMBL Data
                       Library, September 1995
GENETICS
   #gene        SGD:ORC3; OAF1
      ##cross-references SGD:S0003927; MIPS:YLL004w
   #map_position 12L
SUMMARY         #length 616  #molecular-weight 72076  #checksum 76

  Query Match             6.7%;  Score 107;  DB 2;  Length 616;
  Best Local Similarity  25.7%;  Pred. No. 2.07e+00;
  Matches    26;  Conservative   33;  Mismatches  32;  Indels  10;  Gaps  10;

Db     197 KDLAMVFNFKDVDSINFNTLDNFIILLKSAFKYDHVKISLIFNINTNLSNIEKNLRQSTI 256
           :::|: :|:    | :  :||: |:||::|| |   :|:  :: | :: | :  |:   |
Qy      36 NEVALKLNYHPA-SEKVQALDEKILLLRPAFQYSD-NIAKEYE-N-KFKN-QTALKVEQI 90

Db     257 RLLKRNYHKLDV-SSNKG-FKYGNQIFQSFLD-TVDGKLNL 294
           | :: |  : | ||:|  | ::    :::|  :::| : |
Qy      91 -LQNQGYKVISVDSSDKDDFSFAQKK-EGYLAVAMNGEIVL 129


RESULT   12
ENTRY           E70135     #type complete
TITLE           flagellar protein (flbC) homolog - Lyme disease spirochete
ORGANISM        #formal_name Borrelia burgdorferi #common_name Lyme disease
                  spirochete
DATE            13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
                  05-Jun-1998
ACCESSIONS      E70135
REFERENCE       A70100
   #authors     Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
                  Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
                  Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
                  J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
                  Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
                  Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
                  J.; Utterback, T.; Watthey, L.; McDonald, L.; Artiach, P.;
                  Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
                  K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
   #journal     Nature (1997) 390:580-586
   #title       Genomic sequence of a Lyme disease spirochaete, Borrelia
                  burgdorferi.
   #cross-references MUID:98065943
   #accession   E70135
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-394 ##label KLE
      ##cross-references GB:AE001137; GB:AE000783; NID:g2688160; PID:g2688181;
                        TIGR:BB0285
      ##experimental_source strain B31
SUMMARY         #length 394  #molecular-weight 45073  #checksum 5653

  Query Match             6.6%;  Score 106;  DB 2;  Length 394;
  Best Local Similarity  19.7%;  Pred. No. 2.63e+00;
  Matches    24;  Conservative   44;  Mismatches  48;  Indels   6;  Gaps   6;

Db     249 DFVSSLKPEWNLKINKNIVDKA-KVV-LKSNNTGEIKLVLKPKELGSIRINLNLDSNNNL 306
           :: : :| :  ||::  : : : ||: : |::  :: :  | || | : : :| :
Qy      74 EYENKFKNQTALKVEQILQNQGYKVISVDSSDKDDFSFAQK-KE-GYLAVAMNGEIVLRP 131

Db     307 LGKIVVDNQNVK-MLFDQNMHSLNKMLGESGF-NASLNLFLAGENLNSFTGNFKDDSKDQ 364
             |  ::: :   :||  :  :: :|  :|| : ::   ::||:|:||| :: :   ::
Qy     132 DPKRTIQKKSEPGLLFSTGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQE 191

Db     365 NF 366
           :|
Qy     192 KF 193


RESULT   13
ENTRY           S73374     #type complete
TITLE           phenylalanine--tRNA ligase (EC 6.1.1.20) beta chain pheT -
                  Mycoplasma pneumoniae (ATCC 29342) (SGC3)
ALTERNATE_NAMES hypothetical protein C09_orf805; phenylalanyl-tRNA synthetase
                  beta chain pheT
ORGANISM        #formal_name Mycoplasma pneumoniae
   #variety     ATCC 29342
DATE            26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change
                  16-Jul-1999
ACCESSIONS      S73374
REFERENCE       S73327
   #authors     Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li,
```

```
                  B.C.; Herrmann, R.
   #journal       Nucleic Acids Res. (1996) 24:4420-4449
   #title         Complete sequence analysis of the genome of the bacterium
                    Mycoplasma pneumoniae.
   #cross-references MUID:97105885
   #accession     S73374
      ##status          preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-805 ##label HIM
      ##cross-references EMBL:AE000006; GB:U00089; NID:g1673695;
                         PIDN:AAB95696.1; PID:g1673698
      ##note           the nucleotide sequence was submitted to the EMBL Data
                         Library, November 1996
GENETICS
   #gene          pheT
   #genetic_code SGC3
CLASSIFICATION    #superfamily phenylalanine--tRNA ligase beta chain
KEYWORDS          aminoacyl-tRNA synthetase; ligase; protein biosynthesis
SUMMARY           #length 805  #molecular-weight 91713  #checksum 4320

  Query Match             6.6%;  Score 105;  DB 2;  Length 805;
  Best Local Similarity  28.4%;  Pred. No. 3.33e+00;
  Matches    23;  Conservative   21;  Mismatches  33;  Indels   4;  Gaps   4;

Db     506 ALQKLRTKLQTLGFHNVITYQLISPERARNFNLFGLSNLWEIKNP-LSNERSVLRVGLID 564
           :|:|:   |   || :|   : :| |   :| :  ||:| :| :  |   :|    ||:
Qy     150 GLDKMEGVLIPAGFIKVTILEPMSGESLDSFTM-DLSEL-DIQEKFLKTTHSSHSGGLVS 207

Db     565 SLLRVIQK-NAAYKNKLGNIF 584
           ::::  :: | | |: | :||
Qy     208 TMVKGTDNSNDAIKSALNKIF 228


RESULT   14
ENTRY             S31262     #type fragment
TITLE             TyB protein - yeast (Saccharomyces cerevisiae)
                    retrotransposon Ty4 (fragment)
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              18-Jun-1993 #sequence_revision 25-Apr-1997 #text_change
                    25-Apr-1997
ACCESSIONS        S31262
REFERENCE         S31261
   #authors       Janetzky, B.; Lehle, L.
   #journal       J. Biol. Chem. (1992) 267:19798-19805
   #title         Ty4, a new retrotransposon from Saccharomyces cerevisiae,
                    flanked by tau-elements.
   #cross-references MUID:93015829
   #accession     S31262
      ##molecule_type DNA
      ##residues       1-1465 ##label JAN
      ##cross-references EMBL:S46865
GENETICS
   #mobile_element retrotransposon Ty4
SUMMARY           #length 1465  #checksum 1908

  Query Match             6.6%;  Score 105;  DB 2;  Length 1465;
  Best Local Similarity  25.6%;  Pred. No. 3.33e+00;
  Matches    21;  Conservative   28;  Mismatches  29;  Indels   4;  Gaps   4;

Db     833 NLKELKVTDKNVPTDNGTNVSPRLEQNIEASGSPVQTVNKSAFLNKEFSSLNMKRKRKRH 892
           :| || : :| : | :::: :  :   : :: :: ::: ||| ||| |::: |    |:
Qy     183 DLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDNSNDAI-KSA-LNKIFANI-MQEIDKKL 239

Db     893 DKNNSLTSYELERDKKRSKKNR 914
             :| | ||: :    ::|:||
Qy     240 TQKN-LESYQKDAKELKGKRNR 260


RESULT   15
ENTRY             S56894     #type complete
TITLE             TyB protein - yeast (Saccharomyces cerevisiae)
                    retrotransposon Ty4.JL
ALTERNATE_NAMES   protein J0780; protein YJL113w
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change
                    14-Nov-1997
ACCESSIONS        S56894
REFERENCE         S56891
   #authors       Cziepluch, C.; Kordes, E.; Pujol, A.; Jauniaux, J.C.
   #submission    submitted to the Protein Sequence Database, September 1995
   #accession     S56894
      ##molecule_type DNA
      ##residues       1-1803 ##label CZI
      ##cross-references EMBL:Z49389; MIPS:YJL113w
GENETICS
   #map_position 10L
   #mobile_element retrotransposon Ty4.JL
SUMMARY           #length 1803  #molecular-weight 207708  #checksum 3450

  Query Match             6.6%;  Score 105;  DB 2;  Length 1803;
  Best Local Similarity  25.6%;  Pred. No. 3.33e+00;
  Matches    21;  Conservative   28;  Mismatches  29;  Indels   4;  Gaps   4;

Db    1171 NLKELKVTDKNVPTDNGTNVSPRLEQNIEASGSPVQTVNKSAFLNKEFSSLNMKRKRKRH 1230
           :| || : :| : | :::: :  :   : :: :: ::: ||| ||| |::: |    |:
Qy     183 DLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDNSNDAI-KSA-LNKIFANI-MQEIDKKL 239

Db    1231 DKNNSLTSYELERDKKRSKKNR 1252
             :| | ||: :    ::|:||
Qy     240 TQKN-LESYQKDAKELKGKRNR 260


Search completed: Tue Sep 26 16:15:47 2000
Job time : 19 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 26 16:14:59 2000; MasPar time 9.09 Seconds
606.971 Million cell updates/sec
Tabular output not generated.

Title: >US-09-308-435-2
Description: (28-260) from US09308435.pep
Perfect Score: 1594
Sequence: 1 CSPHIIETNEVALKLNYHPA.....QKNLESYQKDAKELKGKRNR 233

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 32.420; Variance 175.543; scale 0.185

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1594	100.0	260	1	W60156	Helicobacter pylori 29	4.75e-113
2	1594	100.0	260	1	W07449	Helicobacter pylori su	4.75e-113
3	1594	100.0	461	1	W74466	Adhesin/V.cholerae tox	4.75e-113
4	1589	99.7	260	1	W60157	Helicobacter pylori 29	1.18e-112
5	1589	99.7	260	1	W07450	Helicobacter pylori su	1.18e-112
6	1552	97.4	260	1	Y11000	H. pylori ORF 11ap2071	9.72e-110
7	1545	96.9	268	1	W20973	H. pylori derived prot	3.46e-109
8	302	18.9	97	1	W24663	H. pylori derived prot	5.42e-13
9	302	18.9	97	1	W20454	H. pylori derived prot	5.42e-13
10	230	14.4	249	1	W55445	H. pylori ORF 01ae1101	8.55e-08
11	230	14.4	261	1	W55332	H. pylori ORF hp3e1107	8.55e-08
12	222	13.9	249	1	W98332	H. pylori GHPO 1615 pr	3.15e-07
13	171	10.7	282	1	Y10951	H. pylori ORF 04ee1110	1.10e-03
14	171	10.7	285	1	W20938	H. pylori secreted or	1.10e-03
15	118	7.4	147	1	W20468	H. pylori secreted or	3.33e+00
16	109	6.8	549	1	W21638	Grapevine leafroll vir	1.21e+01
17	107	6.7	615	1	R77276	ORC3 subunit of yeast	1.60e+01
18	107	6.7	615	1	W22226	S. cerevisiae origin o	1.60e+01
19	102	6.4	552	1	W90172	Human heart muscle spe	3.23e+01
20	100	6.3	751	1	W13491	Helicobacter ClpB (N-t	4.25e+01
21	100	6.3	1254	1	W24575	Merozoite apical-end p	4.25e+01
22	100	6.3	1254	1	R07503	Merozite apical-end-lo	4.25e+01
23	99	6.2	1285	1	P93089	Pasteurella multocida	4.88e+01
24	97	6.1	119	1	W49813	Amino acid sequence of	6.42e+01
25	97	6.1	222	1	R39267	Humanised C4G1 Ig heav	6.42e+01
26	97	6.1	222	1	W49817	Fragment of humanised	6.42e+01
27	97	6.1	235	1	R39268	Humanised C4G1 Ig heav	6.42e+01
28	97	6.1	235	1	W49818	Amino acid sequence of	6.42e+01
29	97	6.1	449	1	R43339	Completely humanised C	6.42e+01
30	97	6.1	449	1	W49816	Amino acid sequence of	6.42e+01
31	98	6.1	764	1	W98549	H. pylori GHPO 1604 pr	5.60e+01
32	97	6.1	914	1	W24800	Spinocerebellar ataxia	6.42e+01
33	97	6.1	1312	1	W33807	Human ataxin-2.	6.42e+01
34	97	6.1	1313	1	W60213	Spinocerebellar ataxia	6.42e+01
35	96	6.0	246	1	W95018	Petunia MADS box trans	7.36e+01
36	95	6.0	1140	1	R72386	XAP-1, part of the DNA	8.43e+01
37	96	6.0	2039	1	W56322	Haemophilus paragallin	7.36e+01
38	96	6.0	2042	1	W56319	Haemophilus paragallin	7.36e+01
39	94	5.9	119	1	W49814	Amino acid sequence of	9.65e+01
40	94	5.9	138	1	R39266	Mouse C4G1 Ig heavy-ch	9.65e+01
41	94	5.9	476	1	W02616	Wheat adenylosuccinate	9.65e+01
42	94	5.9	1084	1	R71015	Human neuronal calcium	9.65e+01
43	94	5.9	1091	1	R33553	Sequence of the alpha	9.65e+01
44	94	5.9	1091	1	W63145	Human calcium channel	9.65e+01
45	94	5.9	1091	1	R71011	Human neuronal calcium	9.65e+01

ALIGNMENTS

RESULT 1
ID W60156 standard; Protein; 260 AA.
AC W60156;
DT 28-SEP-1998 (first entry)
DE Helicobacter pylori 29 kDa surface-exposed antigen.
KW Cell surface antigen; adhesin; infection; vaccine; drug delivery.
OS Helicobacter pylori.
PN WO9822135-A1.
PD 28-MAY-1998.
PF 18-NOV-1997; SE1927.
PR 22-NOV-1996; SE-004296.
PA (ASTR) ASTRA AB.
PI Berglindh T, Lofroth J;
DR WPI; 98-312174/27.
DR N-PSDB; V34642.
PT Aggregate of negatively charged lipid and Helicobacter pylori
PT antigen - useful in therapeutic or preventative vaccines against
PT both forms of bacterium, to generate systemic and mucosal antibody
PT responses
PS Claim 6; Page 28-30; 51pp; English.
CC This polypeptide comprises a 29 kDa surface antigen which is
CC expressed on the surface of both dividing (bacillary) and resting
CC (colloid) forms of Helicobacter pylori. The antigen gives rise to
CC both systemic and local (mucosal) production of antibodies. It is
CC an adhesin that is conserved in all tested strains of H. pylori,
CC and is also a putative virulence factor. The invention relates to
CC a new pharmaceutical composition comprising a lipid aggregate of a
CC negatively charged lipid or lipid mixture and at least one
CC antigenic, native or recombinant antigen of H. pylori, preferably
CC the surface-exposed 29 kDa antigen, or a nucleic acid (see V34642)
CC coding for such an antigen. The new formulations, which are useful
CC as vaccine formulations, elicit a protective immune response
CC against H. pylori infections, and are suitable for therapeutic and
CC prophylactic use. An adequate response is produced even without
CC adjuvant. A second 29 kDa antigen (see W60157) is provided that
CC differs only at residue 222 (Arg for Ser).
SQ Sequence 260 AA;

Query Match 100.0%; Score 1594; DB 1; Length 260;
Best Local Similarity 100.0%; Pred. No. 4.75e-113;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db     28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db     88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    2
ID    W07449 standard; Protein; 260 AA.
AC    W07449;
DT    04-MAR-1997  (first entry)
DE    Helicobacter pylori surface-exposed 29 kDa antigen.
KW    Antigen; adhesin; immunogen; vaccine; immunisation; ulcer;
KW    therapy; diagnosis.
OS    Helicobacter pylori strain CCUG 17874 (NTCC 11637).
FH    Key             Location/Qualifiers
FT    peptide         1. .27
FT                    /label= Sig_peptide
FT    cleavage_site   25. .28
FT                    /note= "consensus recognition site for signal
FT                    peptidase II in prolipoproteins"
FT    protein         28. .260
FT                    /label= Mat_protein
FT                    /note= "either the uncleaved or mature protein can
FT                    be used in applns. of the invention"
PN    WO9638475-A1.
PD    05-DEC-1996.
PF    03-JUN-1996; SE0727.
PR    01-JUN-1995; SE-002007.
PR    21-MAR-1996; SE-001085.
PA    (ASTR ) ASTRA AB.
PI    Bolin I,  Svennerholm A;
DR    WPI; 97-034307/03.
DR    N-PSDB; T43834.
PT    Helicobacter pylori surface exposed antigen - useful for treatment,
PT    prophylaxis or diagnosis of mammalian infections
PS    Claim 2; Page 34-35; 49pp; English.
CC    A 29 kDa antigen (W07449) of Helicobacter pylori is an adhesin
CC    expressed on the surface of resting and dividing forms of the
CC    bacterium.  It is conserved in all tested strains of H. pylori,
CC    and gives rise to both systemic and local (mucosal) prodn. of
CC    antibodies.  Its amino acid sequence was deduced from a cDNA clone
CC    (T43834); a sequence (W07550) deduced from a second cDNA clone
CC    (T43835) differs only at position 222 (Arg for Ser).  Recombinant
CC    antigen can be produced in transformed host cells.  It is useful
CC    for the diagnosis and treatment of H. pylori infections and for
CC    use as a vaccine.
SQ    Sequence   260 AA;

  Query Match          100.0%;  Score 1594;  DB 1;  Length 260;
  Best Local Similarity 100.0%;  Pred. No. 4.75e-113;
  Matches   233;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    3
ID    W74466 standard; Protein; 461 AA.
AC    W74466;
DT    18-MAY-1999  (first entry)
DE    Adhesin/V.cholerae toxin A2 and B subunit fusion protein.
KW    Fusion gene; H. pylori infection; adhesin gene; V. cholerae toxin;
KW    A2 subunit; B subunit; diagnosis; vaccine; gastritis; gastric ulcer;
KW    duodenal ulcer.
OS    Chimeric - Helicobacter pylori.
OS    Chimeric - Vibrio cholerae.
PN    WO9853082-A1.
PD    26-NOV-1998.
PF    21-MAY-1997; 000091.
PR    21-MAY-1997; WO-KR0091.
PA    (DAEW-) DAEWOONG PHARM CO LTD.
PI    Kim BO, Lee BK, Park SK, Yoon SW, Yu YH;
DR    WPI; 98-434925/37.
DR    N-PSDB; X21955.
PT    Fusion gene consisting of the Helicobacter pylori adhesin gene
PT    ligated to the A2 and B subunit gene of Vibrio cholerae toxin -
PT    useful in the diagnosis and treatment of H. pylori
PS    Claim 2; Page 20-21; 32pp; English.
CC    This sequence represents the fusion protein of the invention, between the
CC    H. pylori adhesin gene, and the V. cholerae toxin A2 and B subunit genes.
CC    This protein is used in the diagnosis of and as a vaccine against
CC    H. pylori, which is the main cause of gastritis and gastric/duodenal
CC    ulcers. It may also be used in the production of an anti-H. pylori
CC    antibody. Vectors containing the DNA sequence and E. coli strains
CC    containing it form the basis of a process to produce the protein, which
CC    consists of culturing the microorganism and obtaining the protein. The
CC    protein is an effective vaccine against H. pylori due to its excellent
CC    immunogenicity for H. pylori, stability within the stomach environment,
CC    and its ability to penetrate through the stomach mucous membrane to
CC    stimulate sIgA production.
CC    Note: This sequence was indexed from WO9853082, which is the first
CC    major country equivalent to KR97059278.
SQ    Sequence   461 AA;

  Query Match          100.0%;  Score 1594;  DB 1;  Length 461;
  Best Local Similarity 100.0%;  Pred. No. 4.75e-113;
  Matches   233;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      39 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 98
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      99 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 158
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     159 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 218
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     219 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 271
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    4
ID    W60157 standard; Protein; 260 AA.
AC    W60157;
DT    28-SEP-1998  (first entry)
DE    Helicobacter pylori 29 kDa surface-exposed antigen.
KW    Cell surface antigen; adhesin; infection; vaccine; drug delivery.
OS    Helicobacter pylori.
PN    WO9822135-A1.
PD    28-MAY-1998.
PF    18-NOV-1997; SE1927.
PR    22-NOV-1996; SE-004296.
PA    (ASTR ) ASTRA AB.
```

```
PI   Berglindh T, Lofroth J;
DR   WPI; 98-312174/27.
DR   N-PSDB; V34643.
PT   Aggregate of negatively charged lipid and Helicobacter pylori
PT   antigen - useful in therapeutic or preventative vaccines against
PT   both forms of bacterium, to generate systemic and mucosal antibody
PT   responses
PS   Disclosure; Page 35-36; 51pp; English.
CC   This polypeptide comprises a 29 kDa surface antigen which is
CC   expressed on the surface of both dividing (bacillary) and resting
CC   (colloid) forms of Helicobacter pylori.  The antigen gives rise to
CC   both systemic and local (mucosal) production of antibodies.  It is
CC   an adhesin that is conserved in all tested strains of H. pylori,
CC   and is also a putative virulence factor.  The invention relates to
CC   a new pharmaceutical composition comprising a lipid aggregate of a
CC   negatively charged lipid or lipid mixture and at least one
CC   antigenic, native or recombinant antigen of H. pylori, preferably
CC   the surface-exposed 29 kDa antigen, or a nucleic acid (see V34643)
CC   coding for such an antigen.  The new formulations, which are useful
CC   as vaccine formulations, elicit a protective immune response
CC   against H. pylori infections, and are suitable for therapeutic and
CC   prophylactic use.  An adequate response is produced even without
CC   adjuvant.  A second, claimed 29 kDa antigen (see W60156) is
CC   provided that differs only at residue 222 (Ser for Arg).
SQ   Sequence   260 AA;

  Query Match           99.7%;  Score 1589;  DB 1;  Length 260;
  Best Local Similarity  99.6%;  Pred. No. 1.18e-112;
  Matches   232;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
           |||||||||||||| |||||||||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    5
ID   W07450 standard; Protein; 260 AA.
AC   W07450;
DT   04-MAR-1997  (first entry)
DE   Helicobacter pylori surface-exposed 29 kDa antigen.
KW   Antigen; adhesin; immunogen; vaccine; immunisation; ulcer;
KW   therapy; diagnosis.
OS   Helicobacter pylori strain CCUG 17874 (NTCC 11637).
FH   Key             Location/Qualifiers
FT   peptide         1. .27
FT                   /label= Sig_peptide
FT   cleavage_site   25. .28
FT                   /note= "consensus recognition site for signal
FT                   peptidase II in prolipoproteins"
FT   protein         28. .260
FT                   /label= Mat_protein
FT                   /note= "either the uncleaved or mature protein can
FT                   be used in applns. of the invention"
PN   WO9638475-A1.
PD   05-DEC-1996.
PF   03-JUN-1996; SE0727.
PR   01-JUN-1995; SE-002007.
PR   21-MAR-1996; SE-001085.
PA   (ASTR ) ASTRA AB.
PI   Bolin I,  Svennerholm A;
DR   WPI; 97-034307/03.
DR   N-PSDB; T43835.
PT   Helicobacter pylori surface exposed antigen - useful for treatment,
PT   prophylaxis or diagnosis of mammalian infections
PS   Claim 2; Page 37-38; 49pp; English.
CC   A 29 kDa antigen (W07450) of Helicobacter pylori is an adhesin
CC   expressed on the surface of resting and dividing forms of the
CC   bacterium.  It is conserved in all tested strains of H. pylori,
CC   and gives rise to both systemic and local (mucosal) prodn. of
CC   antibodies.  Its amino acid sequence was deduced from a cDNA clone
CC   (T43835); a sequence (W07549) deduced from a second cDNA clone
CC   (T43834) differs only at position 222 (Ser for Arg).  Recombinant
CC   antigen can be produced in transformed host cells.  It is useful
CC   for the diagnosis and treatment of H. pylori infections and for
CC   use as a vaccine.
SQ   Sequence   260 AA;

  Query Match           99.7%;  Score 1589;  DB 1;  Length 260;
  Best Local Similarity  99.6%;  Pred. No. 1.18e-112;
  Matches   232;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
           |||||||||||||| |||||||||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    6
ID   Y11000 standard; Protein; 260 AA.
AC   Y11000;
DT   08-JUN-1999  (first entry)
DE   H. pylori ORF 11ap20714_4797137_f3_45 cell envelope protein.
KW   Vaccine; probe; diagnostic; ORF; cell envelope protein;
KW   secreted protein; cellular protein.
OS   Helicobacter pylori.
PN   WO9818323-A1.
PD   07-MAY-1998.
PF   28-OCT-1997; U19575.
PR   14-JUL-1997; US-891928.
PR   28-OCT-1996; US-739150.
PR   06-DEC-1996; US-759739.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 98-271811/24.
DR   N-PSDB; X30467.
PT   Helicobacter pylori nucleic acids and proteins - used to develop
PT   products for the detection, prevention and treatment of H. pylori
PT   infections
PS   Claims 27, 31; Page 207-208; 279pp; English.
CC   Recombinant or substantially pure preparations of H. pylori polypeptides
CC   are disclosed, together with the nucleic acids encoding them. In all,
CC   73 ORFs are shown. The proteins are variously cell envelope proteins,
CC   secreted proteins or other cellular proteins. Vaccines containing the
CC   nucleic acids or proteins are claimed, as are probes containing at least
CC   8 nucleotides from the nucleic acid sequences. The vaccines are useful
CC   for treating or reducing the risk of H. pylori infections, and the
CC   probes can be used diagnostically for detecting the presence of
CC   Helicobacter in a sample. The products are also of use in screening
CC   for compounds having the ability to interfere with the H. pylori life
CC   cycle or to inhibit H. pylori infection.
SQ   Sequence   260 AA;

  Query Match           97.4%;  Score 1552;  DB 1;  Length 260;
```

Best Local Similarity 96.1%; Pred. No. 9.72e-110;
Matches 224; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

```
Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTTLKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EEILQNQGYKVINVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           |:||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMERVLIPAGFVKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           |||||||| ||||||:|||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFASIMQEMDKKLTQRNLESYQKDAKELKNKRNR 260
           |||||||||||||||||:||||:||||||:|||||||||||| ||||
Qy     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
```

RESULT 7
ID W20973 standard; Protein; 268 AA.
XC W20973;
DT 21-JUL-1997 (first entry)
DE H. pylori derived protein.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW diagnosis.
OS Helicobacter pylori.
PN WO9640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaerd BL;
DR WPI; 97-052306/05.
DR N-PSDB; T68226.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Disclosure; Page 1363-64; 1481pp; English.
CC This sequence represents a H. pylori protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 268 AA;

Query Match 96.9%; Score 1545; DB 1; Length 268;
Best Local Similarity 95.7%; Pred. No. 3.46e-109;
Matches 223; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

```
Db      36 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTTLKV 95
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      96 EEILQNQGYKVINVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 155
           |:||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     156 STGLDKMERVLIPAGFVKVTILKPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 215
           |||||||| ||||||:||||| ||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     216 TMVKGTDNSNDAIKSALNKIFASIMQEMDKKLTQRNLESYQKDAKELKNKRNR 268
           |||||||||||||||||||||:||||:||||||:|||||||||||| ||||
Qy     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
```

RESULT 8
ID W24663 standard; Protein; 97 AA.
AC W24663;
DT 12-AUG-1997 (first entry)
DE H. pylori derived protein.
KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;
KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
KW activator; inhibitor; bacterial life cycle; vaccine; immunise;
KW detection; antisense; inhibition.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT misc_difference 94
FT /note= "encoded by GAK"
PN WO9719098-A1.
PD 29-MAY-1997.
PF 15-NOV-1996; U18542.
PR 17-NOV-1995; US-561469.
PA (ASTR) ASTRA AB.
PI Smith DH;
DR WPI; 97-298052/27.
DR N-PSDB; T77481.
PT Helicobacter pylori nucleic acid sequences and related proteins -
PT used for diagnostics and therapeutics
PS Claim 18; Page 178; 235pp; English.
CC This sequence represents an H. pylori derived protein, no further
CC details are given in the specification.
CC Helicobacter pylori has been strongly linked to chronic gastritis and
CC duodenal ulcer disease. The nucleic acid sequences of the invention
CC are used to evaluate compounds, especially activators or inhibitors of
CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid
CC sequence. The nucleic acid sequences, and corresponding proteins, are
CC also useful for generating vaccines for immunising subjects against H.
CC pylori or for use in detecting the presence of Helicobacter species in
CC a sample. Antisense nucleic acid sequences of these sequences are
CC used to inhibit expression of a gene from Helicobacter species. H.
CC pylori whole genomic DNA was isolated and nebulised to a median size of
CC 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique
CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are
CC complementary to the BstXI-cut pMPX vectors, while the overhang is not
CC self-complementary. Therefore the linkers will not concatemerise nor
CC will the cut vector re-ligate itself easily. The linker-adapter inserts
CC were ligated to each of the 20 pMPX vectors to construct a series of
CC shotgun subclone libraries. The purified DNA samples were then
CC sequenced.
CC Note: The ORF/protein reference number for this sequence was obtained
CC from the related specification, WO9640893.
SQ Sequence 97 AA;

Query Match 18.9%; Score 302; DB 1; Length 97;
Best Local Similarity 97.7%; Pred. No. 5.42e-13;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

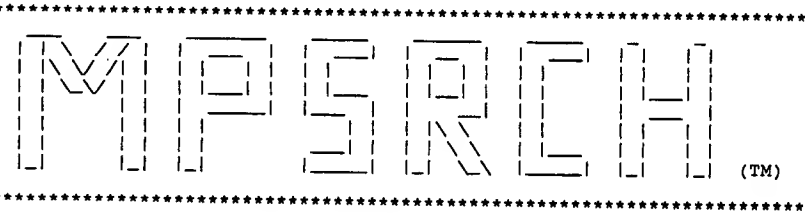
```
Db      53 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSXNI 96
           |||||||||||||||||||||||||||||||||||||||||||| ||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNI 71
```

RESULT 9
ID W20454 standard; protein; 97 AA.
AC W20454;
DT 29-JUL-1997 (first entry)
DE H. pylori derived protein.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS Helicobacter pylori.

```
FH   Key             Location/Qualifiers
FT   misc_difference 94
FT                   /label= Unknown
FT                   /note= "encoded by GAK"
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   N-PSDB; T67801.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Disclosure; Page 625; 1481pp; English.
CC   The present sequence is a H. pylori derived protein, no further details
CC   are given in the specification.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
SQ   Sequence   97 AA;

  Query Match           18.9%;  Score 302;  DB 1;  Length 97;
  Best Local Similarity  97.7%;  Pred. No. 5.42e-13;
  Matches    43;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      53 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSXNI 96
           |||||||||||||||||||||||||||||||||||||||||||| ||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNI 71


RESULT   10
 )   W55445 standard; Protein; 249 AA.
AC   W55445;
DT   24-JUN-1998  (first entry)
DE   H. pylori ORF 01ael1010_40688_c2_38 cell envelope OMP.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS   Helicobacter pylori.
PN   WO9737044-A1.
PD   09-OCT-1997.
PF   27-MAR-1997; U05223.
PR   06-DEC-1996; US-761318.
PR   29-MAR-1996; US-625811.
PR   02-APR-1996; US-758731.
PR   25-OCT-1996; US-736905.
PR   28-OCT-1996; US-738859.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 97-503122/46.
DR   N-PSDB; V24854.
PT   Helicobacter pylori nucleic acid sequences and encoded
PT   polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT   infection and for diagnosis of H. pylori infection
PS   Claims 14,80; Page 653; 1145pp; English.
CC   This sequence is a H. pylori cell envelope outer membrane protein (OMP).
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors. The
CC   DNA and probes derived from it may be used for the identification of
CC   H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC   acid sequences complementary to the DNA act as antisense sequences and
CC   can be used to prevent the translation of H. pylori mRNA. Antibodies
CC   against the protein can be used in immunoassays to evaluate the abundance
CC   and distribution of H. pylori-specific antigens. The genomic sequence of
CC   H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC   by mechanically shearing the bacterial DNA. The sequences were analysed
CC   for ORF of at least 180 nucleotides, and the predicted coding regions
CC   defined by computer evaluation. To identify likely H. pylori antigens for
CC   vaccine development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences of
CC   interest, particular regions can be isolated from H. pylori by PCR
CC   amplification for recombinant polypeptide production, e.g. in E. coli
CC   hosts.
SQ   Sequence   249 AA;

  Query Match           14.4%;  Score 230;  DB 1;  Length 249;
  Best Local Similarity  25.1%;  Pred. No. 8.55e-08;
  Matches    57;  Conservative   74;  Mismatches  74;  Indels  22;  Gaps  21;

Db      40 VELHFHY-PIKGKQEPKNNHLVVLIDPKIEANKVIPENYQKEFEKSLFLQLSNFLERKGY 98
           | | ::| |   | :: ::  ::|: | :: :  |: :|:: | :   | :  :|:  ||
Qy      38 VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db      99 SV-S-QFKDVSEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E-EKVID 147
             | :  |  :::   |  : | : |:|:: :  :    ::  |:: | |   :|:
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     148 -M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTFVHRIK 203
            : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: | | :
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TMV-KGT 213

Db     204 ETDHDRAIKKIMNQAYHKVMAHITKELSKKHMERYEKVSSEMK-KRK 249
           :  :| |||  :|  : ::|  | | |: |::| |:| : |:| ||:
Qy     214 DNSND-AIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259


RESULT   11
ID   W55332 standard; Protein; 261 AA.
AC   W55332;
DT   15-JUN-1998  (first entry)
DE   H. pylori ORF hp3el1075orf3 protein.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   Misc_difference 261
FT                   /note= "the nucleotides encoding this amino acid
FT                         are not given in the specification"
PN   WO9737044-A1.
PD   09-OCT-1997.
PF   27-MAR-1997; U05223.
PR   06-DEC-1996; US-761318.
PR   29-MAR-1996; US-625811.
PR   02-APR-1996; US-758731.
PR   25-OCT-1996; US-736905.
PR   28-OCT-1996; US-738859.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 97-503122/46.
DR   N-PSDB; V24741.
PT   Helicobacter pylori nucleic acid sequences and encoded
PT   polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT   infection and for diagnosis of H. pylori infection
PS   Claim 14; Pages 558-559; 1145pp; English.
CC   This sequence is a H. pylori protein of unspecified function.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors. The
CC   DNA and probes derived from it may be used for the identification of
CC   H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
```

```
CC  acid sequences complementary to the DNA act as antisense sequences and
CC  can be used to prevent the translation of H. pylori mRNA. Antibodies
CC  against the protein can be used in immunoassays to evaluate the abundance
CC  and distribution of H. pylori-specific antigens. The genomic sequence of
CC  H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC  by mechanically shearing the bacterial DNA. The sequences were analysed
CC  for ORF of at least 180 nucleotides, and the predicted coding regions
CC  defined by computer evaluation. To identify likely H. pylori antigens for
CC  vaccine development, the amino acid sequences predicted from various ORF
CC  were analysed for significant homology to other known or exported
CC  membrane proteins. Having identified and determined the sequences of
CC  interest, particular regions can be isolated from H. pylori by PCR
CC  amplification for recombinant polypeptide production, e.g. in E. coli
CC  hosts.
SQ  Sequence   261 AA;

  Query Match           14.4%;  Score 230;  DB 1;  Length 261;
  Best Local Similarity  25.1%;  Pred. No. 8.55e-08;
  Matches    57;  Conservative   74;  Mismatches  74;  Indels  22;  Gaps  21;

Db      52 VELHFHY-PIKGKQEPKNNHLVVLIDPKIEANKVIPENYQKEFEKSLFLQLSNFLERKGY 110
           | | ::| |   | :: ::  ::|: | :: :  |: :|:: | :   | :  :|:  ||
Qy      38 VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db     111 SV-S-QFKDVSEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E-EKVID 159
           | | :  |  :::   |  : | : |:|:: :  :    ::  |::  | |   :|:
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     160 -M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTFVHRIK 215
           : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: | | :
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TMV-KGT 213

Db     216 ETDHDRAIKKIMNQAYHKVMAHITKELSKKHMERYEKVSSEMK-KRK 261
           :  :| |||  :|  : ::|  | | |: |::| |:| : |:| ||:
Qy     214 DNSND-AIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259


RESULT   12
ID   W98332 standard; Protein; 249 AA.
AC   W98332;
DT   31-MAR-1999  (first entry)
DE   H. pylori GHPO 1615 protein.
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease.
OS   Helicobacter pylori.
PN   WO9843478-A1.
PD   08-OCT-1998.
PF   01-APR-1998; U06371.
PR   29-JUL-1997; US-902615.
PR   01-APR-1997; US-833457.
PR   24-JUN-1997; US-881227.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI   Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR   WPI; 98-542293/46.
DR   N-PSDB; X14051.
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
PS   Claim 8; Page 455-456; 2054pp; English.
CC   This sequence represents a Helicobacter pylori GHPO protein of the
CC   invention. The polypeptides can be used for preventing or treating
CC   Helicobacter infections, and gastroduodenal diseases associated with
CC   these infections, including acute, chronic, and atrophic gastritis, and
CC   peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC   used for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
SQ   Sequence   249 AA;

  Query Match           13.9%;  Score 222;  DB 1;  Length 249;
  Best Local Similarity  25.0%;  Pred. No. 3.15e-07;
  Matches    58;  Conservative   74;  Mismatches  77;  Indels  23;  Gaps  22;

Db      35 NTGESVELHFHY-PIKGKQEPKNSHLVVLIEPKIEINKVIPESYQKEFEKSLFLQLSSFL 93
           :| | | | ::| |   | :: :   ::|: | :: :  |:  |:: | :   | :  :|
Qy      34 ETNE-VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQIL 91

Db      94 ERKGYSV-S-QFKDASEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E- 142
           :  || | | :  |  :::   |  : | : |:|:: :  :    ::  |::  | |
Qy      92 QNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGL 151

Db     143 EKVID-M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTF 198
           :|:   : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: |
Qy     152 DKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TM 209

Db     199 VHRIKETDHDQAIRKIMNQAYHKVMVHITKELSKKHMEHYEKVSSEMK-KRK 249
           | :  :  :| ||:  :|  : ::|  | | |: |::| |:| : |:| ||:
Qy     210 V-KGTDNSND-AIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259


RESULT   13
ID   Y10951 standard; Protein; 282 AA.
AC   Y10951;
DT   08-JUN-1999  (first entry)
DE   H. pylori ORF 04eel1108_3906963_f1_7 cell envelope protein.
KW   Vaccine; probe; diagnostic; ORF; cell envelope protein;
KW   secreted protein; cellular protein.
OS   Helicobacter pylori.
PN   WO9818323-A1.
PD   07-MAY-1998.
PF   28-OCT-1997; U19575.
PR   14-JUL-1997; US-891928.
PR   28-OCT-1996; US-739150.
PR   06-DEC-1996; US-759739.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 98-271811/24.
DR   N-PSDB; X30418.
PT   Helicobacter pylori nucleic acids and proteins - used to develop
PT   products for the detection, prevention and treatment of H. pylori
PT   infections
PS   Claims 27, 31; Page 151; 279pp; English.
CC   Recombinant or substantially pure preparations of H. pylori polypeptides
CC   are disclosed, together with the nucleic acids encoding them. In all,
CC   73 ORFs are shown. The proteins are variously cell envelope proteins,
CC   secreted proteins or other cellular proteins. Vaccines containing the
CC   nucleic acids or proteins are claimed, as are probes containing at least
CC   8 nucleotides from the nucleic acid sequences. The vaccines are useful
CC   for treating or reducing the risk of H. pylori infections, and the
CC   probes can be used diagnostically for detecting the presence of
CC   Helicobacter in a sample. The products are also of use in screening
CC   for compounds having the ability to interfere with the H. pylori life
CC   cycle or to inhibit H. pylori infection.
SQ   Sequence   282 AA;

  Query Match           10.7%;  Score 171;  DB 1;  Length 282;
  Best Local Similarity  25.1%;  Pred. No. 1.10e-03;
  Matches    56;  Conservative   58;  Mismatches  91;  Indels  18;  Gaps  14;

Db      64 PLDFNY-PV-HIVQAPQNHHVVGILMPRIQVSDNL-KPYIDKFQDALINQIQTIFEKRGY 120
           :| :|| |    |||  :  :: :| | :| |||: | | :|| :      :: |::::||
Qy      39 ALKLNYHPASEKVQA-LDEKIL-LLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db     121 QVLRFQD-EKA-LNVQDKKKIFSVLDLKGWVGILEDLKMNL--K-DPN---SPNLDTL-- 170
           |:  :  :|  ::  :||  :  : ::| : :  | |  :  | :|    |  || :
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     171 VDQSSGSVWFNFYEPESNRVVHDFAVEVGTFQAITYTYTSTNNASGGFNSSKSVIHENLD 230
           |  ::| :   : || |   :  |::::: :: |   :  | ::| :     :::    |
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSELD-IQEKFLKTTHSSHSGGLVSTMVKGT-D 214

Db     231 KNREDAIHKILNRMYAVVMKKAVTELTKENIAKYRDAIDRMKG 273
           | :|||   ||:::| :|      ||  |:  |:     :||
Qy     215 -NSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKG 256
```

```
RESULT   14
ID   W20938 standard; protein; 285 AA.
AC   W20938;
DT   21-JUL-1997  (first entry)
DE   H. pylori secreted or periplasmic protein, 27zel0351orf5.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS   Helicobacter pylori.
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   N-PSDB; T68191.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Claim 72; Page 1333-1334; 1481pp; English.
CC   The present sequence is a H. pylori secreted or periplasmic protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
SQ   Sequence   285 AA;

  Query Match           10.7%;  Score 171;  DB 1;  Length 285;
  Best Local Similarity  25.1%;  Pred. No. 1.10e-03;
  Matches    56;  Conservative   58;  Mismatches  91;  Indels  18;  Gaps  14;

Db      67 PLDFNY-PV-HIVQAPQNHHVVGILMPRIQVSDNL-KPYIDKFQDALINQIQTIFEKRGY 123
           :| :|| |    |||  :  :: :| | :| |||: | | :|| :     :: |::::||
Qy      39 ALKLNYHPASEKVQA-LDEKIL-LLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db     124 QVLRFQD-EKA-LNVQDKKKIFSVLDLKGWVGILEDLKMNL--K-DPN---SPNLDTL-- 173
            |:  :  :|  ::  :||  :  : ::| : :  | |  :  | :|    |  || :
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     174 VDQSSGSVWFNFYEPESNRVVHDFAVEVGTFQAITYTYTSTNNASGGFNSSKSVIHENLD 233
           |  ::| :   : || |   :  |::::: :: |   :  | ::| :     :::    |
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSELD-IQEKFLKTTHSSHSGGLVSTMVKGT-D 214

Db     234 KNREDAIHKILNRMYAVVMKKAVTELTKENIAKYRDAIDRMKG 276
           | :|||   ||:::| :|      || |:  |:     :||
Qy     215 -NSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKG 256


RESULT   15
ID   W20468 standard; Protein; 147 AA.
AC   W20468;
DT   11-JUL-1997  (first entry)
DE   H. pylori secreted or peripplsmic protein, 3906963.aa.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacterium; life cycle; activator;
KW   bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW   diagnosis.
OS   Helicobacter pylori.
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   N-PSDB; T67633.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Claim 72; Page 638; 1481pp; English.
CC   This sequence represents a H. pylori secreted or periplasmic protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
SQ   Sequence   147 AA;

  Query Match           7.4%;  Score 118;  DB 1;  Length 147;
  Best Local Similarity  28.1%;  Pred. No. 3.33e+00;
  Matches    25;  Conservative   26;  Mismatches  31;  Indels   7;  Gaps   7;

Db      64 PLDFNY-PV-HIVQAPQNHHVVGILMPRIQVSDNL-KPYIDKFQDALINQIQTIFEKRGY 120
           :| :|| |    |||  :  :: :| | :| |||: | | :|| :     :: |::::||
Qy      39 ALKLNYHPASEKVQA-LDEKIL-LLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db     121 QVLRFQD-EKA-LNVQDKKKIFSVLDLKG 147
           |:  :  :|  ::  :||  :  : ::|
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNG 125


Search completed: Tue Sep 26 16:15:11 2000
Job time : 12 secs.
```



fPsrch_pp protein - protein database search, using Smith-Waterman algorithm

:un on: Tue Sep 26 16:20:31 2000; MasPar time 23.59 Seconds
684.688 Million cell updates/sec

abular output not generated.

itle: >US-09-308-435-4
escription: (28-260) from US09308435.pep
erfect Score: 1599
equence: 1 CSPHIIETNEVALKLNYHPA.....QKNLESYQKDAKELKGKRNR 233

coring table: PAM 150
Gap 11

earched: 225878 seqs, 69334122 residues

ost-processing: Minimum Match 0%
Listing first 45 summaries

-tabase: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

-atistics: Mean 46.340; Variance 102.801; scale 0.451

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1559	97.5	260	2	Q9ZL47	NEURAMINYLLACTOSE-BIND	3.46e-262
2	235	14.7	249	2	Q9ZKG5	PUTATIVE PARALOG OF HP	2.97e-19
3	227	14.2	249	2	O25166	PUTATIVE NEURAMINYLLAC	5.53e-18
4	176	11.0	282	2	Q9ZLY2	PUTATIVE PARALOG OF HP	3.63e-10
5	166	10.4	278	2	O25234	HYPOTHETICAL 31.9 KD P	1.05e-08
6	139	8.7	179	9	O48453	COMPLETE NUCLEOTIDE SE	6.35e-05
7	117	7.3	578	5	O97281	PFC0895W PROTEIN.	4.55e-02
8	116	7.3	1027	2	O06471	CEOB.	6.05e-02
9	114	7.1	417	5	O02080	SIMILAR TO CCAAT/ENHAN	1.07e-01
10	109	6.8	549	14	O71192	59 KDA PROTEIN.	4.27e-01
11	107	6.7	9376	2	O85168	SYRINGOMYCIN SYNTHETAS	7.37e-01
12	105	6.6	387	13	O93582	PAIRED TYPE HOMEODOMAI	1.26e+00
13	106	6.6	394	2	Q44769	FLAGELLAR PROTEIN (FLB	9.65e-01
14	105	6.6	911	5	O77338	PFC0565W PROTEIN.	1.26e+00
15	104	6.5	272	5	Q94895	YOLK PROTEIN 1 (FRAGME	1.65e+00
16	104	6.5	402	1	Q58526	GTP-BINDING PROTEIN HF	1.65e+00
17	104	6.5	501	10	O64949	ENDO-1,4-BETA GLUCANAS	1.65e+00
18	104	6.5	512	2	O05176	SUGAR TRANSPORTER.	1.65e+00
19	103	6.4	394	2	Q44919	FTSA.	2.15e+00
20	103	6.4	769	3	O74253	CELLOBIOSE DEHYDROGENA	2.15e+00
21	100	6.3	221	2	O84533	L3 RIBOSOMAL PROTEIN.	4.71e+00
22	101	6.3	248	5	Q9XW20	Y18D10A.11 PROTEIN.	3.63e+00
23	101	6.3	319	2	Q45595	FUNCTION UNKNOWN.	3.63e+00
24	100	6.3	399	11	O35137	ARISTALESS HOMEOBOX-4	4.71e+00
25	101	6.3	583	13	Q91850	SODIUM-CALCIUM EXCHANG	3.63e+00
26	101	6.3	592	1	O26557	SENSORY TRANSDUCTION R	3.63e+00
27	101	6.3	945	5	Q9Y1J9	PUTATIVE VOLTAGE-AND C	3.63e+00
28	100	6.3	1803	3	P87192	CHROMOSOME X READING F	4.71e+00
29	99	6.2	202	2	Q9X1P6	SERINE CYCLE ENZYME, P	6.09e+00
30	99	6.2	235	2	Q47380	INVASIN.	6.09e+00
31	99	6.2	559	2	Q48382	TRANSPOSASE TNIA.	6.09e+00
32	99	6.2	559	12	Q9WTI1	TRANSPOSASE OF IN21.	6.09e+00
33	99	6.2	571	2	Q56381	TNIA.	6.09e+00
34	99	6.2	571	2	O32213	YVGQ PROTEIN.	6.09e+00
35	98	6.1	202	2	O08230	OUTER SURFACE PROTEIN	7.87e+00
36	98	6.1	269	2	Q9XCQ1	TRANSCRIPTIONAL REGULA	7.87e+00
37	97	6.1	277	5	Q94969	YOLK PROTEIN 1 (FRAGME	1.01e+01
38	97	6.1	405	10	O23784	PATATIN-LIKE PROTEIN (	1.01e+01
39	97	6.1	505	5	O01328	F46A8.2 PROTEIN.	1.01e+01
40	98	6.1	559	2	Q9ZEG9	TNIA PROTEIN.	7.87e+00
41	97	6.1	604	5	P91863	F32A7.6 PROTEIN.	1.01e+01
42	97	6.1	877	3	Q04007	HYPOTHETICAL 98.3 KD P	1.01e+01
43	98	6.1	980	3	Q00760	DEVELOPMENTAL REGULATO	7.87e+00
44	97	6.1	1312	4	Q99700	ATAXIN-2.	1.01e+01
45	98	6.1	1790	3	P78621	FH1/FH2 PROTEIN HOMOLO	7.87e+00

ALIGNMENTS

```
RESULT    1
ID   Q9ZL47       PRELIMINARY;      PRT;   260 AA.
AC   Q9ZL47;
DT   01-MAY-1999 (TrEMBLrel..10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR.
GN   HPAA.
OS   Helicobacter pylori J99.
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J99;
RX   MEDLINE; 99120557.
RA   ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA   SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
RA   TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
RA   GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
RA   TRUST T.J.;
RT   "Genomic-sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
DR   EMBL; AE001504; AAD06306.1; -.
SQ   SEQUENCE   260 AA;  29199 MW;  9C0103A4 CRC32;

  Query Match           97.5%;  Score 1559;  DB 2;  Length 260;
  Best Local Similarity 96.1%;  Pred. No. 3.46e-262;
  Matches   224;  Conservative    7;  Mismatches   2;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTTLKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EEILQNQGYKVINVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           |:||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFVKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFASIMQEMDKKLTQRNLESYQKDAKELKNKRNR 260
           |||||||||||||| |||||||:||||:|||||||:|||||||||||||| ||||
```

```
    208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260

SULT    2
   Q9ZKG5      PRELIMINARY;      PRT;   249 AA.
   Q9ZKG5;
   01-MAY-1999 (TrEMBLrel. 10, Created)
   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
   PUTATIVE PARALOG OF HPAA.
   JHP0971.
   Helicobacter pylori J99.
   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
   Helicobacter.
   [1]
   SEQUENCE FROM N.A.
   STRAIN=J99;
   MEDLINE; 99120557.
   ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
   SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
   TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
   GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
   TRUST T.J.;
   "Genomic-sequence comparison of two unrelated isolates of the human
   gastric pathogen Helicobacter pylori.";
   Nature 397:176-180(1999).
   EMBL; AE001526; AAD06549.1; -.
   SEQUENCE   249 AA;  28513 MW;  A3729951 CRC32;

uery Match           14.7%;  Score 235;  DB 2;  Length 249;
est Local Similarity  25.1%;  Pred. No. 2.97e-19;
latches    57;  Conservative   75;  Mismatches  73;  Indels  22;  Gaps  21;

     40 VELHFHY-PIKGKQEPKNNHLVVLIDPKIEANKVIPENYQKEFEKSLFLQLSNFLERKGY 98
        | | ::| |   | :: ::  ::|: | :: :  |: :|:: | :   | :  :|:  ||
     38 VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

     99 SV-S-QFKDVSEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E-EKVID 147
        | | :  |  :::   |  : | : |:|:: : :   ::  |::  | |   :|:
     97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

    148 -M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTFVHRIK 203
        : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: | | :
    157 VLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TMV-KGT 213

    204 ETDHDRAIKKIMNQAYHKVMAHITKELSKKHMERYEKVSSEMK-KRK 249
        :  :| |||: :|  : ::|  | | |: |::| |:| : |:| ||:
    214 DNSND-AIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259


ULT    3
   O25166      PRELIMINARY;      PRT;   249 AA.
   O25166;
   01-JAN-1998 (TrEMBLrel. 05, Created)
   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
   PUTATIVE NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ HOMOLOG (HPAA).
   HP0410.
   Helicobacter pylori (Campylobacter pylori).
   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
   Helicobacter.
   [1]
   SEQUENCE FROM N.A.
   STRAIN=26695;
   MEDLINE; 97394467.
   TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
   FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
   NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
   LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
   MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
   BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
   COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
   HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
   RA   VENTER J.C.;
   RT   "The complete genome sequence of the gastric pathogen Helicobacter
   RT   pylori.";
   RL   Nature 388:539-547(1997).
   DR   EMBL; AE000557; AAD07478.1; -.
   DR   TIGR; HP0410; -.
   KW   Hypothetical protein.
   SQ   SEQUENCE   249 AA;  28349 MW;  6BED3165 CRC32;

     Query Match           14.2%;  Score 227;  DB 2;  Length 249;
     Best Local Similarity  25.0%;  Pred. No. 5.53e-18;
     Matches    58;  Conservative   75;  Mismatches  76;  Indels  23;  Gaps  22;

Db      35 NTGESVELHFHY-PIKGKQEPKNSHLVVLIEPKIEINKVIPESYQKEFEKSLFLQLSSFL 93
           :| | | | ::| |   | :: :   ::|: | :: :  |:  |:: | :   | :  :|
Qy      34 ETNE-VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQIL 91

Db      94 ERKGYSV-S-QFKDASEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E- 142
           :  || | | :  |  :::   |  : | : |:|:: :  :    ::  |::  | |
Qy      92 QNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGL 151

Db     143 EKVID-M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTF 198
           :|:   : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: |
Qy     152 DKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TM 209

Db     199 VHRIKETDHDQAIRKIMNQAYHKVMVHITKELSKKHMEHYEKVSSEMK-KRK 249
           | :  :  :| ||:: :|  : ::|  | | |: |::| |:| : |:| ||:
Qy     210 V-KGTDNSND-AIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259


RESULT    4
ID   Q9ZLY2      PRELIMINARY;      PRT;   282 AA.
AC   Q9ZLY2;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PUTATIVE PARALOG OF HPAA.
GN   JHP0444.
OS   Helicobacter pylori J99.
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J99;
RX   MEDLINE; 99120557.
RA   ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA   SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
RA   TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
RA   GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
RA   TRUST T.J.;
RT   "Genomic-sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
DR   EMBL; AE001479; AAD06033.1; -.
SQ   SEQUENCE   282 AA;  32075 MW;  E982253F CRC32;

     Query Match           11.0%;  Score 176;  DB 2;  Length 282;
     Best Local Similarity  25.1%;  Pred. No. 3.63e-10;
     Matches    56;  Conservative   59;  Mismatches  90;  Indels  18;  Gaps  14;

Db      64 PLDFNY-PV-HIVQAPQNHHVVGILMPRIQVSDNL-KPYIDKFQDALINQIQTIFEKRGY 120
           :| :|| |    |||  :  :: :| | :| |||: | | :|| :     :: |::::||
Qy      39 ALKLNYHPASEKVQA-LDEKIL-LLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db     121 QVLRFQD-EKA-LNVQDKKKIFSVLDLKGWVGILEDLKMNL--K-DPN---SPNLDTL-- 170
            |:  :  :|  ::  :||  :  : ::| : :  | |  :  | :|    |  || :
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     171 VDQSSGSVWFNFYEPESNRVVHDFAVEVGTFQAITYTYTSTNNASGGFNSSKSVIHENLD 230
           |  ::| :   : || |   :  |::::: :: |  :  | ::| :      :::    |
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSELD-IQEKFLKTTHSSHSGGLVSTMVKGT-D 214

Db     231 KNREDAIHKILNRMYAVVMKKAVTELTKENIAKYRDAIDRMKG 273
```

```
     | :||| : ||:::| :|      ||  |:  |:     :||
215 -NSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKG 256


SULT    5
   O25234      PRELIMINARY;      PRT;   278 AA.
   O25234;
   01-JAN-1998 (TrEMBLrel. 05, Created)
   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
   HYPOTHETICAL 31.9 KD PROTEIN.
   HP0492.
   Helicobacter pylori (Campylobacter pylori).
   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
   Helicobacter.
   [1]
   SEQUENCE FROM N.A.
   STRAIN=26695;
   MEDLINE; 97394467.
   TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
   FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
   NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
   LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
   MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
   BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
   COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
   HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
   VENTER J.C.;
   "The complete genome sequence of the gastric pathogen Helicobacter
   pylori.";
   Nature 388:539-547(1997).
   EMBL; AE000564; AAD07568.1; -.
   TIGR; HP0492; -.
   Hypothetical protein.
   SEQUENCE   278 AA;  31947 MW;  79127867 CRC32;

Query Match           10.4%;  Score 166;  DB 2;  Length 278;
Best Local Similarity  25.0%;  Pred. No. 1.05e-08;
Matches    56;  Conservative   62;  Mismatches  84;  Indels  22;  Gaps  18;

     62 PLDFNY-PI-HIVQAPQNHHVVGILTPRIQVSDNL-KPYIDKFQDALINQIQTIFEKRGY 118
        :| :|| |    |||  :  :: :| | :| |||: | | :|| :     :: |:::|| 
     39 ALKLNYHPASEKVQA-LDEKIL-LLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

    119 QVLRFQD-EKA-LNAQDKRKIFSVLDLKGWVGILEDLKMNL--K-DPN---NPNLDTL-- 168
        |:  :  :|  ::  :|:  :  : ::| : :  | |  :  | :|    :  || :
     97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

    169 VDQSSGSVWFNFYEPESNRVVHDFAVEVGTFQAM-TY-TYKHN-NSGGLNSSNSIIHEYL 225
        |  ::| :   : || |   :  |::::: ::    :    |: :|||| | :::
    157 VLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGL-VS-TMVKG-T 213

    226 EKNKEDAIHKILNRMYAVVMKKAVTELTKENIDKYREAIDRMKG 269
        : | :||| : ||:::| :|      ||  |:: |:     :||
    214 D-NSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKG 256


ULT    6
   O48453      PRELIMINARY;      PRT;   179 AA.
   O48453;
   01-JUN-1998 (TrEMBLrel. 06, Created)
   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
   COMPLETE NUCLEOTIDE SEQUENCE.
   Bacteriophage SPP1.
   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
   Lambda phage group.
   [1]
   SEQUENCE FROM N.A.
   ALONSO J.C., LUDER G., STIEGE A.C., CHAI S., WEISE F., TRAUTNER T.A.;
   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
   EMBL; X97918; CAA66554.1; -.
   SEQUENCE   179 AA;  20956 MW;  85120A6C CRC32;

  Query Match           8.7%;  Score 139;  DB 9;  Length 179;
  Best Local Similarity  28.2%;  Pred. No. 6.35e-05;
  Matches    24;  Conservative   27;  Mismatches  31;  Indels   3;  Gaps   3;

Db     88 SLDDIGQALSDMDIDEAFNQAFQALDNSGFFKRAVRAYWKKNDSIIKALESQ-AKTIEDV 146
          ||| :   ||::||:| |  : ::  ::|:    |::  : ||:| :||:   |: ::::
Qy    176 SLDSFTMDLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDNSNDAIKRALNKIFANIMQEI 235

Db    147 EEKELVLRGMEQY-KEIKKIREEMN 170
          : | |  : :| | |: | ::   |
Qy    236 DKK-LTQKNLESYQKDAKELKGKRN 259


RESULT    7
ID   O97281      PRELIMINARY;      PRT;   578 AA.
AC   O97281;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PFC0895W PROTEIN.
GN   PFC0895W.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RA   OLIVER K., BOWMAN S., HARRIS D., LAWSON D., QUAIL M., BARRELL B.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL034559; CAB39027.1; -.
SQ   SEQUENCE   578 AA;  68384 MW;  1AC1C75C CRC32;

  Query Match           7.3%;  Score 117;  DB 5;  Length 578;
  Best Local Similarity  24.7%;  Pred. No. 4.55e-02;
  Matches    20;  Conservative   22;  Mismatches  34;  Indels   5;  Gaps   5;

Db    123 EKTNSLEETLVQINNEYNKCIDMIEKEIHDNIDDKESLKVKNELEKL-C-VIKKEKDEKE 180
          ||  :|:| :: :   :    | | ||  ::: :  :|||   |::   ||  :  :|:
Qy     49 EKVQALDEKILLLRPAFQ-YSDNIAKEYENKFKNQTALKVEQILQNQGYKVISVDSSDKD 107

Db    181 EYEKYKKKMKEKSMEMDMDNE 201
          ::  :  | ||  : : |: |
Qy    108 DFS-FAQK-KEGYLAVAMNGE 126


RESULT    8
ID   O06471      PRELIMINARY;      PRT;  1027 AA.
AC   O06471;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CEOB.
GN   CEOB.
OS   Burkholderia cepacia (Pseudomonas cepacia).
OC   Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC   Burkholderia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K61-3;
RA   CHARRON R.L., BARRY J.J., BURNS J.L.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U97042; AAB58161.1; -.
DR   PFAM; PF00873; ACR_tran; 1.
DR   PRINTS; PR00702; ACRIFLAVINRP.
SQ   SEQUENCE   1027 AA;  110930 MW;  7C8108F2 CRC32;

  Query Match           7.3%;  Score 116;  DB 2;  Length 1027;
  Best Local Similarity  15.2%;  Pred. No. 6.05e-02;
  Matches    22;  Conservative   62;  Mismatches  53;  Indels   8;  Gaps   8;

Db    405 IGIVVDDAIVVVENVERNIENGMNARQATYKA-MQEVSGPIIAIALTLVAVFVPLAFMSG 463
          ::: ::  ||:  :  | |:: ::     : : :: : | :|: ::  |::: |::   :
Qy    119 LAVAMNGEIVLRPDPKRTIQK-KSEPGLLFSTGLDKMEGVLIPAGFIKVTILEPMS-GES 176
```

```
    464 LTGQFYKQFA-MTIAISTVISAFNSLTLSPALSAILLKGHGDKEDWLTRVMNRVLGGFFR 522
        | : | :::: : |      : : :| : | :| : ::||  : :| : | :|:::: : :
    177 LDS-FTMDLSELDIQ-EKFLKTTHS-SHSGGLVSTMVKGTDNSNDAIKRALNKIFANIMQ 233

    523 GFNKVF-HRGAENYGRGVRGVLSRK 546
          ::| : ::  |:| :  : : :::
    234 EIDKKLTQKNLESYQKDAKELKGKR 258


SULT    9
   O02080      PRELIMINARY;      PRT;   417 AA.
   O02080;
   01-JUL-1997 (TrEMBLrel. 04, Created)
   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
   SIMILAR TO CCAAT/ENHANCER-BINDING PROTEIN.
   C48E7.3.
   Caenorhabditis elegans.
   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
   [1]
   SEQUENCE FROM N.A.
   STRAIN=BRISTOL N2;
   MEDLINE; 94150718.
   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
   elegans.";
   Nature 368:32-38(1994).
   [2]
   SEQUENCE FROM N.A.
   STRAIN=BRISTOL N2;
   WAMSLEY P., KRAMER J.;
   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
   [3]
   SEQUENCE FROM N.A.
   STRAIN=BRISTOL N2;
   WATERSTON R.;
   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
   EMBL; AF000262; AAB52934.1; -.
   HSSP; P03069; 2DGC.
   PFAM; PF00170; bZIP; 1.
   SEQUENCE   417 AA;  48317 MW;  D71C1C08 CRC32;

uery Match             7.1%;  Score 114;  DB 5;  Length 417;
est Local Similarity  24.5%;  Pred. No. 1.07e-01;
atches    40;  Conservative   42;  Mismatches  71;  Indels  10;  Gaps   9;

     43 NDTAEKVDELKKENETLERKVEQLQKE-LSFL--KEMFMAYAKNDGNDGPPPPPPPSSSA 99
        |:|| ||::: :       |: :|: :||   || ::| |:| |:    | |  :
     81 NQTALKVEQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMN-GEIVLRPDPKRTIQK 139

    100 MTETEQEAFADLVAELETQTAEAVKAFSKVFDRME-ETNDD-KEGLSLLKLKN-YEMTAY 156
        ::|     |:  :  :|     |      ::: |  |: |     || | : : :|:
    140 KSEPGL-LFSTGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTH 198

    157 LGELTVLMSKMMKGESITVEPSVKRALKHRVFIEKMKPVEDKM 199
          :    |:| |:||   : : ::|||| :::| : |  :: |:
    199 SSHSGGLVSTMVKGTDNSND-AIKRAL-NKIFANIMQEIDKKL 239


ULT   10
  O71192      PRELIMINARY;      PRT;   549 AA.
  O71192;
```

```
   DT   01-AUG-1998 (TrEMBLrel. 07, Created)
   DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
   DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
   DE   59 KDA PROTEIN.
   OS   grapevine leafroll-associated virus 3.
   OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
   OC   Closterovirus.
   RN   [1]
   RP   SEQUENCE FROM N.A.
   RC   STRAIN=NY1;
   RX   MEDLINE; 98264508.
   RA   LING K.-S., ZHU H.-Y., DRONG R.F., SLIGHTOM J.L., MCFERSON J.R.,
   RA   GONSALVES D.;
   RT   "Nucleotide sequence of the 3'-terminal two-thirds of the grapevine
   RT   leafroll-associated virus-3 genome reveals a typical monopartite
   RT   closterovirus.";
   RL   J. Gen. Virol. 79:1299-1307(1998).
   DR   EMBL; AF037268; AAC40708.1; -.
   SQ   SEQUENCE   549 AA;  59013 MW;  2B128E62 CRC32;

    Query Match             6.8%;  Score 109;  DB 14;  Length 549;
    Best Local Similarity  24.5%;  Pred. No. 4.27e-01;
    Matches    24;  Conservative   30;  Mismatches  41;  Indels   3;  Gaps   3;

Db     444 DPVETDVAQFNLSTDGTVSVIVNGEEV-KNEYLVPGTTNVLDSLVYKSGREDLEAKAIPE 502
           |: : |   |    :| ::| :||| | : :    :   :|:: :| : :|:  ||
Qy     102 DSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEGVLIPA 161

Db     503 YLTTLNILHDKAFTRRNLGNKDKGFSDLRIEENFLKSA 540
             :  : || : : :  :| :  : :|:| |:|:|||::
Qy     162 GFIKVTIL-E-PMSGESLDSFTMDLSELDIQEKFLKTT 197


RESULT   11
ID   O85168      PRELIMINARY;      PRT;  9376 AA.
AC   O85168;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   SYRINGOMYCIN SYNTHETASE.
GN   SYRE.
OS   Pseudomonas syringae (pv. syringae).
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC   Pseudomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   GUENZI E., GRANDI G.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF047828; AAC80285.1; -.
DR   PROSITE; PS00455; AMP_BINDING; 8.
DR   PFAM; PF00501; AMP-binding; 8.
DR   PFAM; PF00668; DUF4; 9.
DR   PFAM; PF00550; pp-binding; 9.
DR   PFAM; PF00975; Thioesterase; 1.
DR   PRINTS; PR00154; AMPBINDING.
SQ   SEQUENCE   9376 AA;  1029826 MW;  E4B2CD95 CRC32;

  Query Match             6.7%;  Score 107;  DB 2;  Length 9376;
  Best Local Similarity  24.4%;  Pred. No. 7.37e-01;
  Matches    29;  Conservative   33;  Mismatches  51;  Indels   6;  Gaps   6;

Db    4322 MQGDDGQPVQRISPADTGFNLQMHDLQG-LADAE-EKLQAL-A-SEESLQSFDLQQGPL- 4376
           :: |  : :|: | :   |:  :  ::| |  |    |:  | : | |||:|| :: : |
Qy     129 LRPDPKRTIQKKSEPGLLFSTGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELD 188

Db    4377 IRGRLIRMAEDHHVLLLTLHHIVSDGWSVDVLTRELSALYAAFSQDQDDPLAPLELQ-Y 4434
           |: :::: :   |   |    : :   | | : | |: ::| : |: |  |:  :|: |
Qy     189 IQEKFLKTTHSSHSGGLVSTMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESY 247


RESULT   12
ID   O93582      PRELIMINARY;      PRT;   387 AA.
AC   O93582;
```

```
   01-NOV-1998 (TrEMBLrel. 08, Created)
   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
   PAIRED TYPE HOMEODOMAIN PROTEIN.
   ALX-4.
   Gallus gallus (Chicken).
   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
   Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
   [1]
   SEQUENCE FROM N.A.
   TAKAHASHI M., TAMURA K., BUSCHER D., MASUYA H., YONEI-TAMURA S.,
   MATSUMOTO K., NAITOH-MATSUO M., TAKEUCHI J., OGURA K., SHIROISHI T.,
   OGURA T., IZPISHUA BELMONTE J.C.;
   "Strong's luxoid (lstD), a 16 base pair deletion in the homeobox of
   Alx-4.";
   Development 0:0-0(1998).
   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
   EMBL; AF092538; AAC61772.1; -.
   HSSP; P06601; 1FJL.
   PROSITE; PS00027; HOMEOBOX_1; 1.
   PFAM; PF00046; homeobox; 1.
   Homeobox; DNA-binding; Nuclear protein.
   SEQUENCE   387 AA;  41738 MW;  C862A1E0 CRC32;

Query Match             6.6%;  Score 105;  DB 13;  Length 387;
Best Local Similarity  25.0%;  Pred. No. 1.26e+00;
Matches    20;  Conservative   20;  Mismatches  38;  Indels   2;  Gaps   2;

    149 DPSGMDSSYLSVKEAGVKVPQDRASTDLPSPMDKADSESNKGKKRRNRTTFTSYQLEELE 208
        |:  || | | : |  :|  :   |  | | | |:  :|| : ||     |::  ::|::
    178 DSFTMDLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDNSNDAIKRALNKIFANI-MQEID 236

    209 KVFQKTHYPDVYAREQLAMR 228
        | :   :  : | ::   ::
    237 KKLTQKNL-ESYQKDAKELK 255


SULT   13
   Q44769      PRELIMINARY;      PRT;   394 AA.
   Q44769;
   01-NOV-1996 (TrEMBLrel. 01, Created)
   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
   FLAGELLAR PROTEIN (FLBC) (ORF20).
   BB0285.
   Borrelia burgdorferi (Lyme disease spirochete).
   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
   [1]
   SEQUENCE FROM N.A.
   MEDLINE; 98065943.
   FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
   LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
   DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
   PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
   VAN-VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
   UTTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
   GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
   SMITH H.O., VENTER J.C.;
   "Genomic sequence of a Lyme disease spirochaete, Borrelia
   burgdorferi.";
   Nature 390:580-586(1997).
   [2]
   SEQUENCE FROM N.A.
   FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
   LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
   DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
   PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
   VAN-VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
   UTTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
   GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
   SMITH H.O., VENTER J.C.;
   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
   [3]
   SEQUENCE OF 3-394 FROM N.A.
   STRAIN=B31;
   DUNN J.J., BUTLER-LOFFREDO L., KIELECZAWA J., MEDALLE J., LUFT B.J.;
   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
   EMBL; AE001137; AAC66663.1; -.
   EMBL; U43739; AAA85608.1; -.
   HSSP; P13466; 1KSR.
   TIGR; BB0285; -.
   Flagella.
   SEQUENCE   394 AA;  45073 MW;  5BF8FA60 CRC32;

  Query Match             6.6%;  Score 106;  DB 2;  Length 394;
  Best Local Similarity  19.7%;  Pred. No. 9.65e-01;
  Matches    24;  Conservative   44;  Mismatches  48;  Indels   6;  Gaps   6;

Db    249 DFVSSLKPEWNLKINKNIVDKA-KVV-LKSNNTGEIKLVLKPKELGSIRINLNLDSNNNL 306
          :: : :|.:  ||::  : : : ||: : |::  :: :  | || | : : :| :
Qy     74 EYENKFKNQTALKVEQILQNQGYKVISVDSSDKDDFSFAQK-KE-GYLAVAMNGEIVLRP 131

Db    307 LGKIVVDNQNVK-MLFDQNMHSLNKMLGESGF-NASLNLFLAGENLNSFTGNFKDDSKDQ 364
            |  ::: :   :||   :  :: :|  :|| : ::   ::||:|:|:|| :: :   ::
Qy    132 DPKRTIQKKSEPGLLFSTGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQE 191

Db    365 NF 366
          :|
Qy    192 KF 193


RESULT   14
ID   O77338      PRELIMINARY;      PRT;   911 AA.
AC   O77338;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PFC0565W PROTEIN.
GN   PFC0565W.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RA   HAMLIN N., LAWSON D., BARRELL B.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL008970; CAA15593.1; -.
SQ   SEQUENCE   911 AA;  108563 MW;  5043B736 CRC32;

  Query Match             6.6%;  Score 105;  DB 5;  Length 911;
  Best Local Similarity  19.8%;  Pred. No. 1.26e+00;
  Matches    25;  Conservative   43;  Mismatches  53;  Indels   5;  Gaps   5;

Db    703 NEDLLFDENNRVYTNIRKSDICIYIKEIKNNNINLNKVDKKMIFYLLKEKKNIIFIVNKI 762
          ::|: | ::: |  : ::  |  ::   :  |  :| :  ::|    :|:: ::|   :
Qy    106 KDDFSFAQKKEGYLAVAMNGE-IVLRPDPKRTIQ-KKSEPGLLFSTGLDKMEGVLIPAGF 163

Db    763 DLILTNFEKKRNDFLQEFS-NLFN-DIPIIFLNTKNNTHINTLLNKIIHIHKMNNVIIST 820
            : :| :| :  : |: |: :| : ||   ||:| :::|   |:: ::     :|  |
Qy    164 -IKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDNSNDAIKR 222

Db    821 SLLNLF 826
          :| ::|
Qy    223 ALNKIF 228


RESULT   15
ID   Q94895      PRELIMINARY;      PRT;   272 AA.
AC   Q94895;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   YOLK PROTEIN 1 (FRAGMENT).
GN   YP1.
OS   Drosophila mimica (Fruit fly) (Idiomyia mimica).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
```

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
[1]
SEQUENCE FROM N.A.
MEDLINE; 96120851.
KAMBYSELLIS M.P., HO K.F., CRADDOCK E.M., PIANO F., PARISI M.,
COHEN J.;
"Pattern of ecological shifts in the diversification of Hawaiian
Drosophila inferred from a molecular phylogeny.";
Curr. Biol. 5:1129-1139(1995).
[2]
SEQUENCE FROM N.A.
HO K.F.;
Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
EMBL; U51875; AAC47246.1; -.
FLYBASE; FBgn0016257; Dmic\Yp1.
PFAM; PF00151; lipase; 1.
NON_TER 1 1
NON_TER 272 272
SEQUENCE 272 AA; 30187 MW; DF3EFC7C CRC32;

Query Match 6.5%; Score 104; DB 5; Length 272;
Best Local Similarity 29.1%; Pred. No. 1.65e+00;
Matches 16; Conservative 15; Mismatches 23; Indels 1; Gaps 1;

116 VTIIVTGLPKTSETVKKANSKLVNAYMQRYNLKQQQQQGQKYNKDSSE-ENNRQR 169
206 VSTMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260

arch completed: Tue Sep 26 16:20:56 2000
b time : 25 secs.

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 26 16:16:34 2000; MasPar time 23.61 Seconds
684.174 Million cell updates/sec

Tabular output not generated.

Title: >US-09-308-435-2
Description: (28-260) from US09308435.pep
Perfect Score: 1594
Sequence: 1 CSPHIIETNEVALKLNYHPA.....QKNLESYQKDAKELKGKRNR 233

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 46.291; Variance 102.838; scale 0.450

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1564	98.1	260	2	Q9ZL47	NEURAMINYLLACTOSE-BIND	8.59e-263
2	230	14.4	249	2	Q9ZKG5	PUTATIVE PARALOG OF HP	1.92e-18
3	222	13.9	249	2	O25166	PUTATIVE NEURAMINYLLAC	3.51e-17
4	171	10.7	282	2	Q9ZLY2	PUTATIVE PARALOG OF HP	1.99e-09
5	161	10.1	278	2	O25234	HYPOTHETICAL 31.9 KD P	5.54e-08
6	134	8.4	179	9	O48453	COMPLETE NUCLEOTIDE SE	2.97e-04
7	117	7.3	578	5	O97281	PFC0895W PROTEIN.	4.54e-02
8	109	6.8	549	14	O71192	59 KDA PROTEIN.	4.25e-01
9	109	6.8	911	5	O77338	PFC0565W PROTEIN.	4.25e-01
10	105	6.6	248	5	Q9XW20	Y18D10A.11 PROTEIN.	1.26e+00
11	106	6.6	394	2	Q44769	FLAGELLAR PROTEIN (FLB	9.60e-01
12	106	6.6	512	2	O05176	SUGAR TRANSPORTER.	9.60e-01
13	106	6.6	1027	2	O06471	CEOB.	9.60e-01
14	105	6.6	1803	3	P87192	CHROMOSOME X READING F	1.26e+00
15	103	6.5	394	2	Q44919	FTSA.	2.14e+00
16	104	6.5	417	5	O02080	SIMILAR TO CCAAT/ENHAN	1.64e+00
17	104	6.5	501	10	O64949	ENDO-1,4-BETA GLUCANAS	1.64e+00
18	103	6.5	769	3	O74253	CELLOBIOSE DEHYDROGENA	2.14e+00
19	102	6.4	202	2	O08230	OUTER SURFACE PROTEIN	2.78e+00
20	102	6.4	235	2	Q47380	INVASIN.	2.78e+00
21	102	6.4	559	12	Q9WTI1	TRANSPOSASE OF IN21.	2.78e+00
22	102	6.4	559	2	Q48382	TRANSPOSASE TNIA.	2.78e+00
23	102	6.4	571	2	Q56381	TNIA.	2.78e+00
24	100	6.3	221	2	O84533	L3 RIBOSOMAL PROTEIN.	4.68e+00
25	101	6.3	319	2	Q45595	FUNCTION UNKNOWN.	3.61e+00
26	101	6.3	559	2	Q9ZEG9	TNIA PROTEIN.	3.61e+00
27	101	6.3	571	2	O32213	YVGQ PROTEIN.	3.61e+00
28	101	6.3	592	1	O26557	SENSORY TRANSDUCTION R	3.61e+00
29	100	6.3	856	2	Q9ZMH1	HEAT SHOCK PROTEIN.	4.68e+00
30	101	6.3	945	5	Q9Y1J9	PUTATIVE VOLTAGE-AND C	3.61e+00
31	101	6.3	1000	9	O80181	PUTATIVE MINOR STRUCTU	3.61e+00
32	99	6.2	272	5	Q94895	YOLK PROTEIN 1 (FRAGME	6.05e+00
33	99	6.2	405	10	O23784	PATATIN-LIKE PROTEIN (	6.05e+00
34	97	6.1	604	5	P91863	F32A7.6 PROTEIN.	1.01e+01
35	98	6.1	845	10	Q9ZVG0	PUTATIVE DNA RECOMBINA	7.82e+00
36	97	6.1	914	4	Q99493	SCA2 PROTEIN (FRAGMENT	1.01e+01
37	98	6.1	980	3	Q00760	DEVELOPMENTAL REGULATO	7.82e+00
38	97	6.1	1312	4	Q99700	ATAXIN-2.	1.01e+01
39	98	6.1	1790	3	P78621	FH1/FH2 PROTEIN HOMOLO	7.82e+00
40	97	6.1	9376	2	O85168	SYRINGOMYCIN SYNTHETAS	1.01e+01
41	96	6.0	110	8	Q9XJY9	MATURASE K (FRAGMENT).	1.30e+01
42	96	6.0	231	1	Q58888	HYPOTHETICAL PROTEIN M	1.30e+01
43	96	6.0	252	2	Q49544	VAA SURFACE LIPOPROTEI	1.30e+01
44	96	6.0	331	5	Q21938	R11D1.10 PROTEIN.	1.30e+01
45	96	6.0	768	3	O42729	CELLOBIOSE DEHYDROGENA	1.30e+01

ALIGNMENTS

RESULT 1
ID Q9ZL47 PRELIMINARY; PRT; 260 AA.
AC Q9ZL47;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR.
GN HPAA.
OS Helicobacter pylori J99.
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J99;
RX MEDLINE; 99120557.
RA ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
RA TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
RA GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
RA TRUST T.J.;
RT "Genomic-sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
DR EMBL; AE001504; AAD06306.1; -.
SQ SEQUENCE 260 AA; 29199 MW; 9C0103A4 CRC32;

Query Match 98.1%; Score 1564; DB 2; Length 260;
Best Local Similarity 96.6%; Pred. No. 8.59e-263;
Matches 225; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

```
Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTTLKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EEILQNQGYKVINVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           |:||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFVKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFASIMQEMDKKLTQRNLESYQKDAKELKNKRNR 260
           |||||||||||||||||||:||||:||||||:||||||||||||| ||||
```

```
Y    208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


ESULT    2
D   Q9ZKG5      PRELIMINARY;      PRT;   249 AA.
C   Q9ZKG5;
T   01-MAY-1999 (TrEMBLrel. 10, Created)
T   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
T   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
E   PUTATIVE PARALOG OF HPAA.
N   JHP0971.
S   Helicobacter pylori J99.
C   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
C   Helicobacter.
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=J99;
X   MEDLINE; 99120557.
A   ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
A   SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
A   TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
A   GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
A   TRUST T.J.;
T   "Genomic-sequence comparison of two unrelated isolates of the human
T   gastric pathogen Helicobacter pylori.";
L   Nature 397:176-180(1999).
R   EMBL; AE001526; AAD06549.1; -.
)   SEQUENCE   249 AA;  28513 MW;  A3729951 CRC32;

Query Match           14.4%;  Score 230;  DB 2;  Length 249;
Best Local Similarity  25.1%;  Pred. No. 1.92e-18;
Matches    57;  Conservative   74;  Mismatches  74;  Indels  22;  Gaps  21;

)     40 VELHFHY-PIKGKQEPKNNHLVVLIDPKIEANKVIPENYQKEFEKSLFLQLSNFLERKGY 98
         | | ::| |   | :: ::  ::|: | :: :  |: :|:: | :   | :  :|:  ||
      38 VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

      99 SV-S-QFKDVSEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E-EKVID 147
          | | :  |  :::   |  : | : |:|:: :  :    ::  |:: | |   :|:
      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

     148 -M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTFVHRIK 203
          : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: | | :
     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TMV-KGT 213

     204 ETDHDRAIKKIMNQAYHKVMAHITKELSKKHMERYEKVSSEMK-KRK 249
         :  :| |||  :|  : ::|  | | |: |::| |:| : |:| ||:
     214 DNSND-AIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259


SULT    3
   O25166      PRELIMINARY;      PRT;   249 AA.
   O25166;
   01-JAN-1998 (TrEMBLrel. 05, Created)
   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
   PUTATIVE NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ HOMOLOG (HPAA).
   HP0410.
   Helicobacter pylori (Campylobacter pylori).
   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
   Helicobacter.
   [1]
   SEQUENCE FROM N.A.
   STRAIN=26695;
   MEDLINE; 97394467.
   TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
   FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
   NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
   LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
   MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
   BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
   COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
   HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
   RA   VENTER J.C.;
   RT   "The complete genome sequence of the gastric pathogen Helicobacter
   RT   pylori.";
   RL   Nature 388:539-547(1997).
   DR   EMBL; AE000557; AAD07478.1; -.
   DR   TIGR; HP0410; -.
   KW   Hypothetical protein.
   SQ   SEQUENCE   249 AA;  28349 MW;  6BED3165 CRC32;

     Query Match           13.9%;  Score 222;  DB 2;  Length 249;
     Best Local Similarity  25.0%;  Pred. No. 3.51e-17;
     Matches    58;  Conservative   74;  Mismatches  77;  Indels  23;  Gaps  22;

   Db     35 NTGESVELHFHY-PIKGKQEPKNSHLVVLIEPKIEINKVIPESYQKEFEKSLFLQLSSFL 93
             :| | | | ::| |   | :: :  ::|: | :: :  |:  |:: | :   | :  :|
   Qy     34 ETNE-VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQIL 91

   Db     94 ERKGYSV-S-QFKDASEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E- 142
             :  || | | :  |  :::   |  : | : |:|:: :  :    ::  |::  | |
   Qy     92 QNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGL 151

   Db    143 EKVID-M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTF 198
             :|:   : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: |
   Qy    152 DKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TM 209

   Db    199 VHRIKETDHDQAIRKIMNQAYHKVMVHITKELSKKHMEHYEKVSSEMK-KRK 249
             | :  :  :| ||:  :|  : ::|  | | |: |::| |:| : |:| ||:
   Qy    210 V-KGTDNSND-AIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259


   RESULT    4
   ID   Q9ZLY2      PRELIMINARY;      PRT;   282 AA.
   AC   Q9ZLY2;
   DT   01-MAY-1999 (TrEMBLrel. 10, Created)
   DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
   DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
   DE   PUTATIVE PARALOG OF HPAA.
   GN   JHP0444.
   OS   Helicobacter pylori J99.
   OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
   OC   Helicobacter.
   RN   [1]
   RP   SEQUENCE FROM N.A.
   RC   STRAIN=J99;
   RX   MEDLINE; 99120557.
   RA   ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
   RA   SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
   RA   TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
   RA   GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
   RA   TRUST T.J.;
   RT   "Genomic-sequence comparison of two unrelated isolates of the human
   RT   gastric pathogen Helicobacter pylori.";
   RL   Nature 397:176-180(1999).
   DR   EMBL; AE001479; AAD06033.1; -.
   SQ   SEQUENCE   282 AA;  32075 MW;  E982253F CRC32;

     Query Match           10.7%;  Score 171;  DB 2;  Length 282;
     Best Local Similarity  25.1%;  Pred. No. 1.99e-09;
     Matches    56;  Conservative   58;  Mismatches  91;  Indels  18;  Gaps  14;

   Db     64 PLDFNY-PV-HIVQAPQNHHVVGILMPRIQVSDNL-KPYIDKFQDALINQIQTIFEKRGY 120
             :| :|| |    |||  :  :: :| | :| |||: | | :|| :     :: |::::||
   Qy     39 ALKLNYHPASEKVQA-LDEKIL-LLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

   Db    121 QVLRFQD-EKA-LNVQDKKKIFSVLDLKGWVGILEDLKMNL--K-DPN---SPNLDTL-- 170
              |:  :  :|  ::  :||  :  : ::| : :  | |  :  | :|    |  || :
   Qy     97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

   Db    171 VDQSSGSVWFNFYEPESNRVVHDFAVEVGTFQAITYTYTSTNNASGGFNSSKSVIHENLD 230
             |  ::| :   : || |   :  |:::: :: |   :  | ::| :     :::    |
   Qy    157 VLIPAGFIKVTILEPMSGESLDSFTMDLSELD-IQEKFLKTTHSSHSGGLVSTMVKGT-D 214

   Db    231 KNREDAIHKILNRMYAVVMKKAVTELTKENIAKYRDAIDRMKG 273
```

```
          | :|||   ||:::|. :|      ||  |:  |:     :||
Qy    215 -NSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKG 256


RESULT    5
ID   O25234      PRELIMINARY;      PRT;   278 AA.
AC   O25234;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 31.9 KD PROTEIN.
GN   HP0492.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695;
RX   MEDLINE; 97394467.
RA   TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA   FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA   NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA   LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA   MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA   BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA   COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA   HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA   VENTER J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
DR   EMBL; AE000564; AAD07568.1; -.
DR   TIGR; HP0492; -.
KW   Hypothetical protein.
SQ   SEQUENCE   278 AA;  31947 MW;  79127867 CRC32;

 Query Match           10.1%;  Score 161;  DB 2;  Length 278;
 Best Local Similarity  25.0%;  Pred. No. 5.54e-08;
 Matches    56;  Conservative   61;  Mismatches  85;  Indels  22;  Gaps  18;

Db     62 PLDFNY-PI-HIVQAPQNHHVVGILTPRIQVSDNL-KPYIDKFQDALINQIQTIFEKRGY 118
          :| :|| |    |||  :  :: :| | :| |||: | | :|| :     :: |::::||
Qy     39 ALKLNYHPASEKVQA-LDEKIL-LLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db    119 QVLRFQD-EKA-LNAQDKRKIFSVLDLKGWVGILEDLKMNL--K-DPN---NPNLDTL-- 168
          |:  :  :|  ::  :|:  :  : ::| : :  | |  :  | :|    :  || :
Qy     97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db    169 VDQSSGSVWFNFYEPESNRVVHDFAVEVGTFQAM-TY-TYKHN-NSGGLNSSNSIIHEYL 225
          |  ::| :  : || |   :  |:::::: ::    :    |: :|||| | :::
Qy    157 VLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGL-VS-TMVKG-T 213

Db    226 EKNKEDAIHKILNRMYAVVMKKAVTELTKENIDKYREAIDRMKG 269
          : | :|||   ||:::| :|      ||  |:: |:     :||
Qy    214 D-NSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKG 256


RESULT    6
ID   O48453      PRELIMINARY;      PRT;   179 AA.
AC   O48453;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   COMPLETE NUCLEOTIDE SEQUENCE.
OS   Bacteriophage SPP1.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC   Lambda phage group.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ALONSO J.C., LUDER G., STIEGE A.C., CHAI S., WEISE F., TRAUTNER T.A.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X97918; CAA66554.1; -.
SQ   SEQUENCE   179 AA;  20956 MW;  85120A6C CRC32;

  Query Match           8.4%;  Score 134;  DB 9;  Length 179;
  Best Local Similarity  28.2%;  Pred. No. 2.97e-04;
  Matches    24;  Conservative   26;  Mismatches  32;  Indels   3;  Gaps   3;

Db     88 SLDDIGQALSDMDIDEAFNQAFQALDNSGFFKRAVRAYWKKNDSIIKALESQ-AKTIEDV 146
          ||| :   ||::||:| |  : ::  ::|:    |::  : ||:|  ||:   |: ::::
Qy    176 SLDSFTMDLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDNSNDAIKSALNKIFANIMQEI 235

Db    147 EEKELVLRGMEQY-KEIKKIREEMN 170
          : | |  : :| | |: | ::   |
Qy    236 DKK-LTQKNLESYQKDAKELKGKRN 259


RESULT    7
ID   O97281      PRELIMINARY;      PRT;   578 AA.
AC   O97281;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PFC0895W PROTEIN.
GN   PFC0895W.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RA   OLIVER K., BOWMAN S., HARRIS D., LAWSON D., QUAIL M., BARRELL B.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL034559; CAB39027.1; -.
SQ   SEQUENCE   578 AA;  68384 MW;  1AC1C75C CRC32;

  Query Match           7.3%;  Score 117;  DB 5;  Length 578;
  Best Local Similarity  24.7%;  Pred. No. 4.54e-02;
  Matches    20;  Conservative   22;  Mismatches  34;  Indels   5;  Gaps   5;

Db    123 EKTNSLEETLVQINNEYNKCIDMIEKEIHDNIDDKESLKVKNELEKL-C-VIKKEKDEKE 180
          ||  :|:| :: :   :    | | ||  ::: :  :|||   |::    ||  :  :|:
Qy     49 EKVQALDEKILLLRPAFQ-YSDNIAKEYENKFKNQTALKVEQILQNQGYKVISVDSSDKD 107

Db    181 EYEKYKKKMKEKSMEMDMDNE 201
          ::  :  | ||  : : |: |
Qy    108 DFS-FAQK-KEGYLAVAMNGE 126


RESULT    8
ID   O71192      PRELIMINARY;      PRT;   549 AA.
AC   O71192;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   59 KDA PROTEIN.
OS   grapevine leafroll-associated virus 3.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NY1;
RX   MEDLINE; 98264508.
RA   LING K.-S., ZHU H.-Y., DRONG R.F., SLIGHTOM J.L., MCFERSON J.R.,
RA   GONSALVES D.;
RT   "Nucleotide sequence of the 3'-terminal two-thirds of the grapevine
RT   leafroll-associated virus-3 genome reveals a typical monopartite
RT   closterovirus.";
RL   J. Gen. Virol. 79:1299-1307(1998).
DR   EMBL; AF037268; AAC40708.1; -.
SQ   SEQUENCE   549 AA;  59013 MW;  2B128E62 CRC32;

  Query Match           6.8%;  Score 109;  DB 14;  Length 549;
  Best Local Similarity  24.5%;  Pred. No. 4.25e-01;
  Matches    24;  Conservative   30;  Mismatches  41;  Indels   3;  Gaps   3;

Db    444 DPVETDVAQFNLSTDGTVSVIVNGEEV-KNEYLVPGTTNVLDSLVYKSGREDLEAKAIPE 502
```

```
     |: : |   |    :| ::| :||| | : :       :   :|:: :| : :|:  ||
   102 DSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEGVLIPA 161

   503 YLTTLNILHDKAFTRRNLGNKDKGFSDLRIEENFLKSA 540
       :  : || : : :  :| :  : :|:| |:|:|||::
   162 GFIKVTIL-E-PMSGESLDSFTMDLSELDIQEKFLKTT 197


ULT    9
   O77338      PRELIMINARY;      PRT;   911 AA.
   O77338;
   01-NOV-1998 (TrEMBLrel. 08, Created)
   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
   PFC0565W PROTEIN.
   PFC0565W.
   Plasmodium falciparum.
   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
   [1]
   SEQUENCE FROM N.A.
   STRAIN=3D7;
   HAMLIN N., LAWSON D., BARRELL B.;
   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
   EMBL; AL008970; CAA15593.1; -.
   SEQUENCE   911 AA;  108563 MW;  5043B736 CRC32;

uery Match             6.8%;  Score 109;  DB 5;  Length 911;
est Local Similarity  19.8%;  Pred. No. 4.25e-01;
atches    25;  Conservative   44;  Mismatches  52;  Indels   5;  Gaps   5;

   703 NEDLLFDENNRVYTNIRKSDICIYIKEIKNNNINLNKVDKKMIFYLLKEKKNIIFIVNKI 762
       ::|: | :::  |  : ::   | ::   :  |  :| :  ::|    :|:: ::|   :
   106 KDDFSFAQKKEGYLAVAMNGE-IVLRPDPKRTIQ-KKSEPGLLFSTGLDKMEGVLIPAGF 163

   763 DLILTNFEKKRNDFLQEFS-NLFN-DIPIIFLNTKNNTHINTLLNKIIHIHKMNNVIIST 820
       : :| :| :  : |: |: :| : ||   ||:| :::|   |:: ::     :|  | :
   164 -IKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDNSNDAIKS 222

   821 SLLNLF 826
       :| ::|
   223 ALNKIF 228


ULT   10
  Q9XW20      PRELIMINARY;      PRT;   248 AA.
  Q9XW20;
  01-NOV-1999 (TrEMBLrel. 12, Created)
  01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
  01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
  Y18D10A.11 PROTEIN.
  Y18D10A.11.
  Caenorhabditis elegans.
  Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
  Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
  [1]
  SEQUENCE FROM N.A.
  MEDLINE; 94150718.
  HARRIS B.;
  "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
  elegans.";
  Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
  [2]
  SEQUENCE FROM N.A.
  MEDLINE; 94150718.
  WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
  BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
  CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
  GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
  JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
  LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
  PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
  SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
  THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
  RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
  RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
  RT   elegans.";
  RL   Nature 368:32-38(1994).
  DR   EMBL; AL034393; CAA22314.1; -.
  SQ   SEQUENCE   248 AA;  27963 MW;  BC3B5F9C CRC32;

    Query Match             6.6%;  Score 105;  DB 5;  Length 248;
    Best Local Similarity  22.0%;  Pred. No. 1.26e+00;
    Matches    28;  Conservative   40;  Mismatches  51;  Indels   8;  Gaps   8;

  Db     74 QKRWEGELASSRAQNEQL-SEEKSQLQKENE-ELLLVL-L-RTEGIVDANKSLSEQLANA 129
            ||: || || :   :  | :: |  :|| :|  ||:   | : ||:: :   :   : ::
  Qy    113 QKK-EGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEGVLIPAGFIKVTILEP 171

  Db    130 QDAEVSMTTLKETVAKLEEENNVLQATWAEERSGLVNELIDTKEKLAKSAQAQTELDESH 189
             :| |: ::   :: |: ::: | :| :   :|||: ::   :: :: |   : |:
  Qy    172 MSGE-SLDSFTMDLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDN-SNDA-IKSALNKIF 228

  Db    190 RDIMTAV 196
             :||  :
  Qy    229 ANIMQEI 235


  RESULT   11
  ID   Q44769      PRELIMINARY;      PRT;   394 AA.
  AC   Q44769;
  DT   01-NOV-1996 (TrEMBLrel. 01, Created)
  DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
  DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
  DE   FLAGELLAR PROTEIN (FLBC) (ORF20).
  GN   BB0285.
  OS   Borrelia burgdorferi (Lyme disease spirochete).
  OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
  RN   [1]
  RP   SEQUENCE FROM N.A.
  RX   MEDLINE; 98065943.
  RA   FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
  RA   LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
  RA   DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
  RA   PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
  RA   VAN-VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
  RA   UTTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
  RA   GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
  RA   SMITH H.O., VENTER J.C.;
  RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
  RT   burgdorferi.";
  RL   Nature 390:580-586(1997).
  RN   [2]
  RP   SEQUENCE FROM N.A.
  RA   FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
  RA   LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
  RA   DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
  RA   PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
  RA   VAN-VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
  RA   UTTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
  RA   GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
  RA   SMITH H.O., VENTER J.C.;
  RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
  RN   [3]
  RP   SEQUENCE OF 3-394 FROM N.A.
  RC   STRAIN=B31;
  RA   DUNN J.J., BUTLER-LOFFREDO L., KIELECZAWA J., MEDALLE J., LUFT B.J.;
  RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
  DR   EMBL; AE001137; AAC66663.1; -.
  DR   EMBL; U43739; AAA85608.1; -.
  DR   HSSP; P13466; 1KSR.
  DR   TIGR; BB0285; -.
  KW   Flagella.
  SQ   SEQUENCE   394 AA;  45073 MW;  5BF8FA60 CRC32;

    Query Match             6.6%;  Score 106;  DB 2;  Length 394;
    Best Local Similarity  19.7%;  Pred. No. 9.60e-01;
```

```
 Matches    24;  Conservative   44;  Mismatches  48;  Indels   6;  Gaps   6;

)b    249 DFVSSLKPEWNLKINKNIVDKA-KVV-LKSNNTGEIKLVLKPKELGSIRINLNLDSNNNL 306
          :: : :| :  ||::  : : : ||: : |::  :: :  | || | : : :| :
)y     74 EYENKFKNQTALKVEQILQNQGYKVISVDSSDKDDFSFAQK-KE-GYLAVAMNGEIVLRP 131

)b    307 LGKIVVDNQNVK-MLFDQNMHSLNKMLGESGF-NASLNLFLAGENLNSFTGNFKDDSKDQ 364
            |  ::: :   :||   :  :: :|  :|| : ::   ::||:|:|||  :: :   ::
)y    132 DPKRTIQKKSEPGLLFSTGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQE 191

)b    365 NF 366
          :|
)y    192 KF 193


:ESULT   12
:D   O05176      PRELIMINARY;      PRT;   512 AA.
:C   O05176;
:T   01-JUL-1997 (TrEMBLrel. 04, Created)
:T   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
:T   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
:E   SUGAR TRANSPORTER.
:N   GGUA.
:S   Agrobacterium tumefaciens.
:C   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
:C   Rhizobiaceae; Agrobacterium.
:N   [1]
:P   SEQUENCE FROM N.A.
:C   STRAIN=C58;
:A   KEMNER J.M., LIANG X., NESTER E.W.;
:L   J. Bacteriol. 179:0-0(0).
:R   EMBL; U91632; AAB51510.1; -.
:R   PFAM; PF00005; ABC_tran; 2.
:W   Sugar transport.
:Q   SEQUENCE   512 AA;  56512 MW;  C6F6AEAC CRC32;

 Query Match            6.6%;  Score 106;  DB 2;  Length 512;
 Best Local Similarity  21.2%;  Pred. No. 9.60e-01;
 Matches    29;  Conservative   45;  Mismatches  54;  Indels   9;  Gaps   8;

)b    295 IAGLMGAGRTEFAMSVFGKSY-GHRITGDVLIDGKPVDVSTVRKAIDAGLAYVTE-DRKH 352
          : :: :: : :|: :    :| :  : |:::  :|    |::|  ::|| : |  |::
)y     98 VISVDSSDKDDFSFAQKKEGYLAVAMNGEIVL--RPDPKRTIQKKSEPGLLFSTGLDKME 155

)      353 LGLVLNDNILHNTTLANLAGVSKASIIDDIKEMKVASDF-RTRLRIRSSGIFQETVN-LS 410
          |:::   ::  | |  ::| |  |:  |: |: :   | :|    :|:|:    |:
)y     156 -GVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVSTMVKGTD 214

)b    411 GGNQQ-KVVLSKWLFSN 426
          :|:  |  |:| :|:|
)y    215 NSNDAIKSALNK-IFAN 230


:ESULT   13
:D   O06471      PRELIMINARY;      PRT;  1027 AA.
:C   O06471;
:T   01-JUL-1997 (TrEMBLrel. 04, Created)
:T   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
:T   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
:E   CEOB.
:N   CEOB.
:S   Burkholderia cepacia (Pseudomonas cepacia).
:C   Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
:C   Burkholderia.
:N   [1]
:P   SEQUENCE FROM N.A.
:C   STRAIN=K61-3;
:A   CHARRON R.L., BARRY J.J., BURNS J.L.;
:L   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
:R   EMBL; U97042; AAB58161.1; -.
:R   PFAM; PF00873; ACR_tran; 1.
:R   PRINTS; PR00702; ACRIFLAVINRP.
:Q   SEQUENCE   1027 AA;  110930 MW;  7C8108F2 CRC32;

  Query Match            6.6%;  Score 106;  DB 2;  Length 1027;
  Best Local Similarity  14.5%;  Pred. No. 9.60e-01;
  Matches    21;  Conservative   62;  Mismatches  54;  Indels   8;  Gaps   8;

Db    405 IGIVVDDAIVVVENVERNIENGMNARQATYKA-MQEVSGPIIAIALTLVAVFVPLAFMSG 463
          ::: ::  ||:  :  | |:: ::     : : :: : | :|: ::  |::: |::   :
Qy    119 LAVAMNGEIVLRPDPKRTIQK-KSEPGLLFSTGLDKMEGVLIPAGFIKVTILEPMS-GES 176

Db    464 LTGQFYKQFA-MTIAISTVISAFNSLTLSPALSAILLKGHGDKEDWLTRVMNRVLGGFFR 522
          | : | :::: : |     : : :| : | :| : ::||  : :| :   :|:::: : :
Qy    177 LDS-FTMDLSELDIQ-EKFLKTTHS-SHSGGLVSTMVKGTDNSNDAIKSALNKIFANIMQ 233

Db    523 GFNKVF-HRGAENYGRGVRGVLSRK 546
            ::| : ::  |:| :  : : :::
Qy    234 EIDKKLTQKNLESYQKDAKELKGKR 258


RESULT   14
ID   P87192      PRELIMINARY;      PRT;  1803 AA.
AC   P87192;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUL-1997 (TrEMBLrel. 04, Last annotation update)
DE   CHROMOSOME X READING FRAME ORF YJL114W.
GN   TY4B.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   CZIEPLUCH C., KORDES E., PUJOL A., JAUNIAUX J.C.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z49389; CAA89409.1; -.
SQ   SEQUENCE   1803 AA;  207693 MW;  9AF6A313 CRC32;

  Query Match            6.6%;  Score 105;  DB 3;  Length 1803;
  Best Local Similarity  25.6%;  Pred. No. 1.26e+00;
  Matches    21;  Conservative   28;  Mismatches  29;  Indels   4;  Gaps   4;

Db   1171 NLKELKVTDKNVPTDNGTNVSPRLEQNIEASGSPVQTVNKSAFLNKEFSSLNMKRKRKRH 1230
          :| || : :| : | :::: :  :   : :: :: ::: ||| ||| |::: |    |:
Qy    183 DLSELDIQEKFLKTTHSSHSGGLVSTMVKGTDNSNDAI-KSA-LNKIFANI-MQEIDKKL 239

Db   1231 DKNNSLTSYELERDKKRSKKNR 1252
            :| | ||: :    ::|:||
Qy    240 TQKN-LESYQKDAKELKGKRNR 260


RESULT   15
ID   Q44919      PRELIMINARY;      PRT;   394 AA.
AC   Q44919;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   FTSA.
GN   FLBC.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=212;
RX   MEDLINE; 97234641.
RA   GE Y., OLD I.G., SAINT GIRONS I., CHARON N.W.;
RT   "Molecular characterization of a large Borrelia burgdorferi motility
RT   operon which is initiated by a consensus sigma70 promoter.";
RL   J. Bacteriol. 179:2289-2299(1997).
DR   EMBL; L76303; AAB51416.1; -.
DR   HSSP; P13466; 1KSR.
```

SEQUENCE 394 AA; 45005 MW; 57F04C42 CRC32;

Query Match 6.5%; Score 103; DB 2; Length 394;
Best Local Similarity 20.9%; Pred. No. 2.14e+00;
Matches 24; Conservative 40; Mismatches 45; Indels 6; Gaps 6;

```
249 DFVSSLKPEWNLKINKNIVDKA-KVV-LKSNNTGEIKLVLKPKELGSIRINLNLDSNNNL 306
    :: : :| :  ||::  : : : ||: : |::  :: :  | || | : : :| :
 74 EYENKFKNQTALKVEQILQNQGYKVISVDSSDKDDFSFAQK-KE-GYLAVAMNGEIVLRP 131

307 LGKIVVDNQNVK-MLFDQNMHSLNKMLGESGF-NASLNLFLAGENLNSFTGDFKD 359
      | ::: :   :||   :  :: :|  :|| : ::   ::|||:|:||| |: :
132 DPKRTIQKKSEPGLLFSTGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSE 186
```

Search completed: Tue Sep 26 16:17:02 2000
Job time : 28 secs.

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

.Psrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 26 16:18:57 2000; MasPar time 8.99 Seconds
613.770 Million cell updates/sec
Tabular output not generated.

Title: >US-09-308-435-4
Description: (28-260) from US09308435.pep
Perfect Score: 1599
Sequence: 1 CSPHIIETNEVALKLNYHPA.....QKNLESYQKDAKELKGKRNR 233

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 32.470; Variance 175.985; scale 0.185

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1599	100.0	260	1	W07450	Helicobacter pylori su	2.57e-113
2	1599	100.0	260	1	W60157	Helicobacter pylori 29	2.57e-113
3	1589	99.4	260	1	W60156	Helicobacter pylori 29	1.58e-112
4	1589	99.4	260	1	W07449	Helicobacter pylori su	1.58e-112
5	1589	99.4	461	1	W74466	Adhesin/V.cholerae tox	1.58e-112
6	1547	96.7	260	1	Y11000	H. pylori ORF 11ap2071	3.20e-109
7	1540	96.3	268	1	W20973	H. pylori derived prot	1.14e-108
8	302	18.9	97	1	W24663	H. pylori derived prot	5.74e-13
9	302	18.9	97	1	W20454	H. pylori derived prot	5.74e-13
10	235	14.7	249	1	W55445	H. pylori ORF 01ael101	3.94e-08
11	235	14.7	261	1	W55332	H. pylori ORF hp3e1107	3.94e-08
12	227	14.2	249	1	W98332	H. pylori GHPO 1615 pr	1.46e-07
13	176	11.0	282	1	Y10951	H. pylori ORF 04ee1110	5.19e-04
14	176	11.0	285	1	W20938	H. pylori secreted or	5.19e-04
15	118	7.4	147	1	W20468	H. pylori secreted or	3.41e+00
16	109	6.8	549	1	W21638	Grapevine leafroll vir	1.23e+01
17	107	6.7	615	1	R77276	ORC3 subunit of yeast	1.63e+01
18	107	6.7	615	1	W22226	S. cerevisiae origin o	1.63e+01
19	103	6.4	2039	1	W56322	Haemophilus paragallin	2.86e+01
20	103	6.4	2042	1	W56319	Haemophilus paragallin	2.86e+01
21	100	6.3	1254	1	R07503	Merozite apical-end-lo	4.33e+01
22	100	6.3	1254	1	W24575	Merozoite apical-end p	4.33e+01
23	99	6.2	1285	1	P93089	Pasteurella multocida	4.97e+01
24	97	6.1	119	1	W49813	Amino acid sequence of	6.53e+01
25	97	6.1	222	1	R39267	Humanised C4G1 Ig heav	6.53e+01
26	97	6.1	222	1	W49817	Fragment of humanised	6.53e+01
27	97	6.1	235	1	R39268	Humanised C4G1 Ig heav	6.53e+01
28	97	6.1	235	1	W49818	Amino acid sequence of	6.53e+01
29	97	6.1	449	1	R43339	Completely humanised C	6.53e+01
30	97	6.1	449	1	W49816	Amino acid sequence of	6.53e+01
31	97	6.1	552	1	W90172	Human heart muscle spe	6.53e+01
32	97	6.1	914	1	W24800	Spinocerebellar ataxia	6.53e+01
33	97	6.1	1312	1	W33807	Human ataxin-2.	6.53e+01
34	97	6.1	1313	1	W60213	Spinocerebellar ataxia	6.53e+01
35	96	6.0	246	1	W95018	Petunia MADS box trans	7.48e+01
36	96	6.0	947	1	R76708	Recombinant cold-resis	7.48e+01
37	96	6.0	947	1	R90924	Corn derived pyruvate	7.48e+01
38	94	5.9	175	1	R62783	Borrelia VS461 antigen	9.81e+01
39	95	5.9	599	1	W73479	Grapevine leafroll vir	8.57e+01
40	95	5.9	751	1	W13491	Helicobacter ClpB (N-t	8.57e+01
41	95	5.9	764	1	W98549	H. pylori GHPO 1604 pr	8.57e+01
42	94	5.9	1084	1	R71015	Human neuronal calcium	9.81e+01
43	94	5.9	1091	1	R71011	Human neuronal calcium	9.81e+01
44	95	5.9	1140	1	R72386	XAP-1, part of the DNA	8.57e+01
45	94	5.9	1392	1	Y06999	Restin protein sequenc	9.81e+01

ALIGNMENTS

RESULT 1
ID W07450 standard; Protein; 260 AA.
AC W07450;
DT 04-MAR-1997 (first entry)
DE Helicobacter pylori surface-exposed 29 kDa antigen.
KW Antigen; adhesin; immunogen; vaccine; immunisation; ulcer;
KW therapy; diagnosis.
OS Helicobacter pylori strain CCUG 17874 (NTCC 11637).
FH Key Location/Qualifiers
FT peptide 1. .27
FT /label= Sig_peptide
FT cleavage_site 25. .28
FT /note= "consensus recognition site for signal
FT peptidase II in prolipoproteins"
FT protein 28. .260
FT /label= Mat_protein
FT /note= "either the uncleaved or mature protein can
FT be used in applns. of the invention"
PN WO9638475-A1.
PD 05-DEC-1996.
PF 03-JUN-1996; SE0727.
PR 01-JUN-1995; SE-002007.
PR 21-MAR-1996; SE-001085.
PA (ASTR) ASTRA AB.
PI Bolin I, Svennerholm A;
DR WPI; 97-034307/03.
DR N-PSDB; T43835.
PT Helicobacter pylori surface exposed antigen - useful for treatment,
PT prophylaxis or diagnosis of mammalian infections
PS Claim 2; Page 37-38; 49pp; English.
CC A 29 kDa antigen (W07450) of Helicobacter pylori is an adhesin
CC expressed on the surface of resting and dividing forms of the
CC bacterium. It is conserved in all tested strains of H. pylori,
CC and gives rise to both systemic and local (mucosal) prodn. of
CC antibodies. Its amino acid sequence was deduced from a cDNA clone
CC (T43835); a sequence (W07549) deduced from a second cDNA clone
CC (T43834) differs only at position 222 (Ser for Arg). Recombinant
CC antigen can be produced in transformed host cells. It is useful
CC for the diagnosis and treatment of H. pylori infections and for
CC use as a vaccine.
SQ Sequence 260 AA;

Query Match 100.0%; Score 1599; DB 1; Length 260;
Best Local Similarity 100.0%; Pred. No. 2.57e-113;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

```
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
```

RESULT 2
ID W60157 standard; Protein; 260 AA.
AC W60157;
DT 28-SEP-1998 (first entry)
DE Helicobacter pylori 29 kDa surface-exposed antigen.
KW Cell surface antigen; adhesin; infection; vaccine; drug delivery.
OS Helicobacter pylori.
WO9822135-A1.
28-MAY-1998.
PF 18-NOV-1997; SE1927.
PR 22-NOV-1996; SE-004296.
PA (ASTR) ASTRA AB.
PI Berglindh T, Lofroth J;
DR WPI; 98-312174/27.
DR N-PSDB; V34643.
PT Aggregate of negatively charged lipid and Helicobacter pylori
PT antigen - useful in therapeutic or preventative vaccines against
PT both forms of bacterium, to generate systemic and mucosal antibody
PT responses
PS Disclosure; Page 35-36; 51pp; English.
CC This polypeptide comprises a 29 kDa surface antigen which is
CC expressed on the surface of both dividing (bacillary) and resting
CC (colloid) forms of Helicobacter pylori. The antigen gives rise to
CC both systemic and local (mucosal) production of antibodies. It is
CC an adhesin that is conserved in all tested strains of H. pylori,
CC and is also a putative virulence factor. The invention relates to
CC a new pharmaceutical composition comprising a lipid aggregate of a
CC negatively charged lipid or lipid mixture and at least one
CC antigenic, native or recombinant antigen of H. pylori, preferably
CC the surface-exposed 29 kDa antigen, or a nucleic acid (see V34643)
CC coding for such an antigen. The new formulations, which are useful
CC as vaccine formulations, elicit a protective immune response
CC against H. pylori infections, and are suitable for therapeutic and
CC prophylactic use. An adequate response is produced even without
CC adjuvant. A second, claimed 29 kDa antigen (see W60156) is
CC provided that differs only at residue 222 (Ser for Arg).
SQ Sequence 260 AA;

Query Match 100.0%; Score 1599; DB 1; Length 260;
Best Local Similarity 100.0%; Pred. No. 2.57e-113;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
```

RESULT 3
ID W60156 standard; Protein; 260 AA.
AC W60156;
DT 28-SEP-1998 (first entry)
DE Helicobacter pylori 29 kDa surface-exposed antigen.
KW Cell surface antigen; adhesin; infection; vaccine; drug delivery.
OS Helicobacter pylori.
PN WO9822135-A1.
PD 28-MAY-1998.
PF 18-NOV-1997; SE1927.
PR 22-NOV-1996; SE-004296.
PA (ASTR) ASTRA AB.
PI Berglindh T, Lofroth J;
DR WPI; 98-312174/27.
DR N-PSDB; V34642.
PT Aggregate of negatively charged lipid and Helicobacter pylori
PT antigen - useful in therapeutic or preventative vaccines against
PT both forms of bacterium, to generate systemic and mucosal antibody
PT responses
PS Claim 6; Page 28-30; 51pp; English.
CC This polypeptide comprises a 29 kDa surface antigen which is
CC expressed on the surface of both dividing (bacillary) and resting
CC (colloid) forms of Helicobacter pylori. The antigen gives rise to
CC both systemic and local (mucosal) production of antibodies. It is
CC an adhesin that is conserved in all tested strains of H. pylori,
CC and is also a putative virulence factor. The invention relates to
CC a new pharmaceutical composition comprising a lipid aggregate of a
CC negatively charged lipid or lipid mixture and at least one
CC antigenic, native or recombinant antigen of H. pylori, preferably
CC the surface-exposed 29 kDa antigen, or a nucleic acid (see V34642)
CC coding for such an antigen. The new formulations, which are useful
CC as vaccine formulations, elicit a protective immune response
CC against H. pylori infections, and are suitable for therapeutic and
CC prophylactic use. An adequate response is produced even without
CC adjuvant. A second 29 kDa antigen (see W60157) is provided that
CC differs only at residue 222 (Arg for Ser).
SQ Sequence 260 AA;

Query Match 99.4%; Score 1589; DB 1; Length 260;
Best Local Similarity 99.6%; Pred. No. 1.58e-112;
Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
```

RESULT 4
ID W07449 standard; Protein; 260 AA.
AC W07449;
DT 04-MAR-1997 (first entry)
DE Helicobacter pylori surface-exposed 29 kDa antigen.
KW Antigen; adhesin; immunogen; vaccine; immunisation; ulcer;
KW therapy; diagnosis.
OS Helicobacter pylori strain CCUG 17874 (NTCC 11637).
FH Key Location/Qualifiers
FT peptide 1. .27
FT /label= Sig_peptide
FT cleavage_site 25. .28

```
FT                   /note= "consensus recognition site for signal
FT                   peptidase II in prolipoproteins"
FT   protein         28. .260
FT                   /label= Mat_protein
FT                   /note= "either the uncleaved or mature protein can
FT                   be used in applns. of the invention"
PN   WO9638475-A1.
PD   05-DEC-1996.
PF   03-JUN-1996; SE0727.
PR   01-JUN-1995; SE-002007.
PR   21-MAR-1996; SE-001085.
PA   (ASTR ) ASTRA AB.
PI   Bolin I,  Svennerholm A;
DR   WPI; 97-034307/03.
DR   N-PSDB; T43834.
PT   Helicobacter pylori surface exposed antigen - useful for treatment,
PT   prophylaxis or diagnosis of mammalian infections
PS   Claim 2; Page 34-35; 49pp; English.
CC   A 29 kDa antigen (W07449) of Helicobacter pylori is an adhesin
CC   expressed on the surface of resting and dividing forms of the
CC   bacterium.  It is conserved in all tested strains of H. pylori,
CC   and gives rise to both systemic and local (mucosal) prodn. of
CC   antibodies.  Its amino acid sequence was deduced from a cDNA clone
CC   (T43834); a sequence (W07550) deduced from a second cDNA clone
CC   (T43835) differs only at position 222 (Arg for Ser).  Recombinant
CC   antigen can be produced in transformed host cells.  It is useful
CC   for the diagnosis and treatment of H. pylori infections and for
CC   use as a vaccine.
SQ   Sequence   260 AA;

  Query Match           99.4%;  Score 1589;  DB 1;  Length 260;
  Best Local Similarity  99.6%;  Pred. No. 1.58e-112;
  Matches   232;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    5
ID   W74466 standard; Protein; 461 AA.
AC   W74466;
DT   18-MAY-1999  (first entry)
DE   Adhesin/V.cholerae toxin A2 and B subunit fusion protein.
KW   Fusion gene; H. pylori infection; adhesin gene; V. cholerae toxin;
KW   A2 subunit; B subunit; diagnosis; vaccine; gastritis; gastric ulcer;
KW   duodenal ulcer.
OS   Chimeric - Helicobacter pylori.
OS   Chimeric - Vibrio cholerae.
PN   WO9853082-A1.
PD   26-NOV-1998.
PF   21-MAY-1997; 000091.
PR   21-MAY-1997; WO-KR0091.
PA   (DAEW-) DAEWOONG PHARM CO LTD.
PI   Kim BO, Lee BK, Park SK, Yoon SW, Yu YH;
DR   WPI; 98-434925/37.
DR   N-PSDB; X21955.
PT   Fusion gene consisting of the Helicobacter pylori adhesin gene
PT   ligated to the A2 and B subunit gene of Vibrio cholerae toxin -
PT   useful in the diagnosis and treatment of H. pylori
PS   Claim 2; Page 20-21; 32pp; English.
CC   This sequence represents the fusion protein of the invention, between the
CC   H. pylori adhesin gene, and the V. cholerae toxin A2 and B subunit genes.
CC   This protein is used in the diagnosis of and as a vaccine against
CC   H. pylori, which is the main cause of gastritis and gastric/duodenal
CC   ulcers. It may also be used in the production of an anti-H. pylori
CC   antibody. Vectors containing the DNA sequence and E. coli strains
CC   containing it form the basis of a process to produce the protein, which
CC   consists of culturing the microorganism and obtaining the protein. The
CC   protein is an effective vaccine against H. pylori due to its excellent
CC   immunogenicity for H. pylori, stability within the stomach environment,
CC   and its ability to penetrate through the stomach mucous membrane to
CC   stimulate sIgA production.
CC   Note: This sequence was indexed from WO9853082, which is the first
CC   major country equivalent to KR97059278.
SQ   Sequence   461 AA;

  Query Match           99.4%;  Score 1589;  DB 1;  Length 461;
  Best Local Similarity  99.6%;  Pred. No. 1.58e-112;
  Matches   232;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      39 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 98
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      99 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 158
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     159 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 218
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     219 TMVKGTDNSNDAIKSALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 271
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    6
ID   Y11000 standard; Protein; 260 AA.
AC   Y11000;
DT   08-JUN-1999  (first entry)
DE   H. pylori ORF 11ap20714_4797137_f3_45 cell envelope protein.
KW   Vaccine; probe; diagnostic; ORF; cell envelope protein;
KW   secreted protein; cellular protein.
OS   Helicobacter pylori.
PN   WO9818323-A1.
PD   07-MAY-1998.
PF   28-OCT-1997; U19575.
PR   14-JUL-1997; US-891928.
PR   28-OCT-1996; US-739150.
PR   06-DEC-1996; US-759739.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 98-271811/24.
DR   N-PSDB; X30467.
PT   Helicobacter pylori nucleic acids and proteins - used to develop
PT   products for the detection, prevention and treatment of H. pylori
PT   infections
PS   Claims 27, 31; Page 207-208; 279pp; English.
CC   Recombinant or substantially pure preparations of H. pylori polypeptides
CC   are disclosed, together with the nucleic acids encoding them. In all,
CC   73 ORFs are shown. The proteins are variously cell envelope proteins,
CC   secreted proteins or other cellular proteins. Vaccines containing the
CC   nucleic acids or proteins are claimed, as are probes containing at least
CC   8 nucleotides from the nucleic acid sequences. The vaccines are useful
CC   for treating or reducing the risk of H. pylori infections, and the
CC   probes can be used diagnostically for detecting the presence of
CC   Helicobacter in a sample. The products are also of use in screening
CC   for compounds having the ability to interfere with the H. pylori life
CC   cycle or to inhibit H. pylori infection.
SQ   Sequence   260 AA;

  Query Match           96.7%;  Score 1547;  DB 1;  Length 260;
```

```
  Best Local Similarity  95.7%;  Pred. No. 3.20e-109;
  Matches   223;  Conservative    7;  Mismatches   3;  Indels   0;  Gaps   0;

Db      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTTLKV 87
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      88 EEILQNQGYKVINVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147
           |:||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     148 STGLDKMERVLIPAGFVKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207
           |||||||| |||||||:|||||||||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     208 TMVKGTDNSNDAIKSALNKIFASIMQEMDKKLTQRNLESYQKDAKELKNKRNR 260
           ||||||||||||| |||||||:||||:||||||:|||||||||||||| ||||
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    7
ID   W20973 standard; Protein; 268 AA.
XX
AC   W20973;
XX
DT   21-JUL-1997  (first entry)
XX
DE   H. pylori derived protein.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacterium; life cycle; activator;
KW   bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW   diagnosis.
XX
OS   Helicobacter pylori.
XX
PN   WO9640893-A1.
XX
PD   19-DEC-1996.
XX
PF   06-JUN-1996; U09122.
XX
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
XX
DR   WPI; 97-052306/05.
DR   N-PSDB; T68226.
XX
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
XX
PS   Disclosure; Page 1363-64; 1481pp; English.
XX
CC   This sequence represents a H. pylori protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
XX
SQ   Sequence   268 AA;

  Query Match           96.3%;  Score 1540;  DB 1;  Length 268;
  Best Local Similarity  95.3%;  Pred. No. 1.14e-108;
  Matches   222;  Conservative    7;  Mismatches   4;  Indels   0;  Gaps   0;

Db      36 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTTLKV 95
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNIAKEYENKFKNQTALKV 87

Db      96 EEILQNQGYKVINVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 155
           |:||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Qy      88 EQILQNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLF 147

Db     156 STGLDKMERVLIPAGFVKVTILKPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 215
           |||||||| |||||||:||||| ||||||||||||||||||||||||||||||||||||||
Qy     148 STGLDKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSELDIQEKFLKTTHSSHSGGLVS 207

Db     216 TMVKGTDNSNDAIKSALNKIFASIMQEMDKKLTQRNLESYQKDAKELKNKRNR 268
           |||||||||||||| |||||||:||||:||||||:|||||||||||||| ||||
Qy     208 TMVKGTDNSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRNR 260


RESULT    8
ID   W24663 standard; Protein; 97 AA.
XX
AC   W24663;
XX
DT   12-AUG-1997  (first entry)
XX
DE   H. pylori derived protein.
XX
KW   Transmembrane; cytoplasmic; cell envelope; flagella; transport;
KW   secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
KW   activator; inhibitor; bacterial life cycle; vaccine; immunise;
KW   detection; antisense; inhibition.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   misc_difference 94
FT                   /note= "encoded by GAK"
XX
PN   WO9719098-A1.
XX
PD   29-MAY-1997.
XX
PF   15-NOV-1996; U18542.
XX
PR   17-NOV-1995; US-561469.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Smith DH;
XX
DR   WPI; 97-298052/27.
DR   N-PSDB; T77481.
XX
PT   Helicobacter pylori nucleic acid sequences and related proteins -
PT   used for diagnostics and therapeutics
XX
PS   Claim 18; Page 178; 235pp; English.
XX
CC   This sequence represents an H. pylori derived protein, no further
CC   details are given in the specification.
CC   Helicobacter pylori has been strongly linked to chronic gastritis and
CC   duodenal ulcer disease.  The nucleic acid sequences of the invention
CC   are used to evaluate compounds, especially activators or inhibitors of
CC   bacterial life cycle, for the ability to bind an H. pylori nucleic acid
CC   sequence.  The nucleic acid sequences, and corresponding proteins, are
CC   also useful for generating vaccines for immunising subjects against H.
CC   pylori or for use in detecting the presence of Helicobacter species in
CC   a sample.  Antisense nucleic acid sequences of these sequences are
CC   used to inhibit expression of a gene from Helicobacter species.  H.
CC   pylori whole genomic DNA was isolated and nebulised to a median size of
CC   2000 bp.  Purified DNA fragments were blunt-ended and ligated to unique
CC   BstXI-linker adapters in 100-1000 fold molar excess.  These linkers are
CC   complementary to the BstXI-cut pMPX vectors, while the overhang is not
CC   self-complementary.  Therefore the linkers will not concatemerise nor
CC   will the cut vector re-ligate itself easily.  The linker-adapter inserts
CC   were ligated to each of the 20 pMPX vectors to construct a series of
CC   shotgun subclone libraries.  The purified DNA samples were then
CC   sequenced.
CC   Note: The ORF/protein reference number for this sequence was obtained
CC   from the related specification, WO9640893.
XX
SQ   Sequence   97 AA;

  Query Match           18.9%;  Score 302;  DB 1;  Length 97;
  Best Local Similarity  97.7%;  Pred. No. 5.74e-13;
  Matches    43;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      53 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSXNI 96
           |||||||||||||||||||||||||||||||||||||||||||| ||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNI 71


RESULT    9
ID   W20454 standard; protein; 97 AA.
XX
AC   W20454;
XX
DT   29-JUL-1997  (first entry)
XX
DE   H. pylori derived protein.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS   Helicobacter pylori.
```

```
FH   Key             Location/Qualifiers
FT   misc_difference 94
FT                   /label= Unknown
FT                   /note= "encoded by GAK"
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   N-PSDB; T67801.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Disclosure; Page 625; 1481pp; English.
CC   The present sequence is a H. pylori derived protein, no further details
CC   are given in the specification.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
SQ   Sequence   97 AA;

  Query Match           18.9%;  Score 302;  DB 1;  Length 97;
  Best Local Similarity  97.7%;  Pred. No. 5.74e-13;
  Matches    43;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      53 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSXNI 96
           |||||||||||||||||||||||||||||||||||||||||||| ||
Qy      28 CSPHIIETNEVALKLNYHPASEKVQALDEKILLLRPAFQYSDNI 71


RESULT   10
    W55445 standard; Protein; 249 AA.
    W55445;
DT   24-JUN-1998  (first entry)
DE   H. pylori ORF 01ae11010_40688_c2_38 cell envelope OMP.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS   Helicobacter pylori.
PN   WO9737044-A1.
PD   09-OCT-1997.
PF   27-MAR-1997; U05223.
PR   06-DEC-1996; US-761318.
PR   29-MAR-1996; US-625811.
PR   02-APR-1996; US-758731.
PR   25-OCT-1996; US-736905.
PR   28-OCT-1996; US-738859.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 97-503122/46.
DR   N-PSDB; V24854.
PT   Helicobacter pylori nucleic acid sequences and encoded
PT   polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT   infection and for diagnosis of H. pylori infection
PS   Claims 14,80; Page 653; 1145pp; English.
CC   This sequence is a H. pylori cell envelope outer membrane protein (OMP).
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors. The
CC   DNA and probes derived from it may be used for the identification of
CC   H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC   acid sequences complementary to the DNA act as antisense sequences and
CC   can be used to prevent the translation of H. pylori mRNA. Antibodies
CC   against the protein can be used in immunoassays to evaluate the abundance
CC   and distribution of H. pylori-specific antigens. The genomic sequence of
CC   H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC   by mechanically shearing the bacterial DNA. The sequences were analysed
CC   for ORF of at least 180 nucleotides, and the predicted coding regions
CC   defined by computer evaluation. To identify likely H. pylori antigens for
CC   vaccine development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences of
CC   interest, particular regions can be isolated from H. pylori by PCR
CC   amplification for recombinant polypeptide production, e.g. in E. coli
CC   hosts.
SQ   Sequence   249 AA;

  Query Match           14.7%;  Score 235;  DB 1;  Length 249;
  Best Local Similarity  25.1%;  Pred. No. 3.94e-08;
  Matches    57;  Conservative   75;  Mismatches  73;  Indels  22;  Gaps  21;

Db      40 VELHFHY-PIKGKQEPKNNHLVVLIDPKIEANKVIPENYQKEFEKSLFLQLSNFLERKGY 98
           | | ::| |   | :: ::  ::|: | :: :  |: :|:: | :   | :  :|:  ||
Qy      38 VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db      99 SV-S-QFKDVSEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E-EKVID 147
           | | :  |  :::   |  : | : |:|:: :   ::  |:: | |   :|:
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     148 -M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTFVHRIK 203
           : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: | | :
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TMV-KGT 213

Db     204 ETDHDRAIKKIMNQAYHKVMAHITKELSKKHMERYEKVSSEMK-KRK 249
           :  :| |||: :|  : ::|  | | |: |::| |:| : |:| ||:
Qy     214 DNSND-AIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259


RESULT   11
ID   W55332 standard; Protein; 261 AA.
AC   W55332;
DT   15-JUN-1998  (first entry)
DE   H. pylori ORF hp3e11075orf3 protein.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   Misc_difference 261
FT                   /note= "the nucleotides encoding this amino acid
FT                         are not given in the specification"
PN   WO9737044-A1.
PD   09-OCT-1997.
PF   27-MAR-1997; U05223.
PR   06-DEC-1996; US-761318.
PR   29-MAR-1996; US-625811.
PR   02-APR-1996; US-758731.
PR   25-OCT-1996; US-736905.
PR   28-OCT-1996; US-738859.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 97-503122/46.
DR   N-PSDB; V24741.
PT   Helicobacter pylori nucleic acid sequences and encoded
PT   polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT   infection and for diagnosis of H. pylori infection
PS   Claim 14; Pages 558-559; 1145pp; English.
CC   This sequence is a H. pylori protein of unspecified function.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors. The
CC   DNA and probes derived from it may be used for the identification of
CC   H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
```

```
CC   acid sequences complementary to the DNA act as antisense sequences and
CC   can be used to prevent the translation of H. pylori mRNA. Antibodies
CC   against the protein can be used in immunoassays to evaluate the abundance
CC   and distribution of H. pylori-specific antigens. The genomic sequence of
CC   H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC   by mechanically shearing the bacterial DNA. The sequences were analysed
CC   for ORF of at least 180 nucleotides, and the predicted coding regions
CC   defined by computer evaluation. To identify likely H. pylori antigens for
CC   vaccine development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences of
CC   interest, particular regions can be isolated from H. pylori by PCR
CC   amplification for recombinant polypeptide production, e.g. in E. coli
CC   hosts.
SQ   Sequence   261 AA;

  Query Match           14.7%;  Score 235;  DB 1;  Length 261;
  Best Local Similarity  25.1%;  Pred. No. 3.94e-08;
  Matches    57;  Conservative   75;  Mismatches  73;  Indels  22;  Gaps  21;

Db      52 VELHFHY-PIKGKQEPKNNHLVVLIDPKIEANKVIPENYQKEFEKSLFLQLSNFLERKGY 110
           | | ::| |   | :: ::  ::|: | :: :  |: :|:: | :   | :  :|:  ||
Qy      38 VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db     111 SV-S-QFKDVSEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E-EKVID 159
           | | :  |  :::   |  : | : |:|:: :  :    ::  |::  | |   :|:
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     160 -M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTFVHRIK 215
            : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: | | :
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TMV-KGT 213

Db     216 ETDHDRAIKKIMNQAYHKVMAHITKELSKKHMERYEKVSSEMK-KRK 261
           :  :| |||: :|  : ::|  | | |: |::| |:| : |:| ||:
Qy     214 DNSND-AIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259


RESULT   12
ID   W98332 standard; Protein; 249 AA.
AC   W98332;
DT   31-MAR-1999  (first entry)
DE   H. pylori GHPO 1615 protein.
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease.
OS   Helicobacter pylori.
PN   WO9843478-A1.
PD   08-OCT-1998.
PF   01-APR-1998; U06371.
PR   29-JUL-1997; US-902615.
PR   01-APR-1997; US-833457.
PR   24-JUN-1997; US-881227.
     (HUMA-) HUMAN GENOME SCI INC.
     (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI   Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR   WPI; 98-542293/46.
DR   N-PSDB; X14051.
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
PS   Claim 8; Page 455-456; 2054pp; English.
CC   This sequence represents a Helicobacter pylori GHPO protein of the
CC   invention. The polypeptides can be used for preventing or treating
CC   Helicobacter infections, and gastroduodenal diseases associated with
CC   these infections, including acute, chronic, and atrophic gastritis, and
CC   peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC   used for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
SQ   Sequence   249 AA;

  Query Match           14.2%;  Score 227;  DB 1;  Length 249;
  Best Local Similarity  25.0%;  Pred. No. 1.46e-07;
  Matches    58;  Conservative   75;  Mismatches  76;  Indels  23;  Gaps  22;

Db      35 NTGESVELHFHY-PIKGKQEPKNSHLVVLIEPKIEINKVIPESYQKEFEKSLFLQLSSFL 93
           :| | | | ::| |   | :: :   ::|: | :: :  |:  |:: | :   | :  :|
Qy      34 ETNE-VALKLNYHPASEKVQALDEK-ILLLRPAFQYSDNIAKEYENKFKNQTALKVEQIL 91

Db      94 ERKGYSV-S-QFKDASEIPQDIKEKALLVLRMDGNVAI-LE-D-IVEE-SDA--L-S-E- 142
           :  || | | :  |  :::   |  : | :|::: :  :    ::  |::  | |
Qy      92 QNQGYKVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGL 151

Db     143 EKVID-M-SSGYLNLNFVEPKSEDIIHSFGIDVSKIKAVIER-VEL-RRTNSGGFVPKTF 198
           :|:   : ::|:::: ::||:| : : || :|:| :  : |: :   : ::|||:|: |
Qy     152 DKMEGVLIPAGFIKVTILEPMSGESLDSFTMDLSEL-DIQEKFLKTTHSSHSGGLVS-TM 209

Db     199 VHRIKETDHDQAIRKIMNQAYHKVMVHITKELSKKHMEHYEKVSSEMK-KRK 249
           | :  :  :| ||:: :|  : ::|  | | |: |::| |:| : |:| ||:
Qy     210 V-KGTDNSND-AIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKGKRN 259


RESULT   13
ID   Y10951 standard; Protein; 282 AA.
AC   Y10951;
DT   08-JUN-1999  (first entry)
DE   H. pylori ORF 04eel1108_3906963_f1_7 cell envelope protein.
KW   Vaccine; probe; diagnostic; ORF; cell envelope protein;
KW   secreted protein; cellular protein.
OS   Helicobacter pylori.
PN   WO9818323-A1.
PD   07-MAY-1998.
PF   28-OCT-1997; U19575.
PR   14-JUL-1997; US-891928.
PR   28-OCT-1996; US-739150.
PR   06-DEC-1996; US-759739.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 98-271811/24.
DR   N-PSDB; X30418.
PT   Helicobacter pylori nucleic acids and proteins - used to develop
PT   products for the detection, prevention and treatment of H. pylori
PT   infections
PS   Claims 27, 31; Page 151; 279pp; English.
CC   Recombinant or substantially pure preparations of H. pylori polypeptides
CC   are disclosed, together with the nucleic acids encoding them. In all,
CC   73 ORFs are shown. The proteins are variously cell envelope proteins,
CC   secreted proteins or other cellular proteins. Vaccines containing the
CC   nucleic acids or proteins are claimed, as are probes containing at least
CC   8 nucleotides from the nucleic acid sequences. The vaccines are useful
CC   for treating or reducing the risk of H. pylori infections, and the
CC   probes can be used diagnostically for detecting the presence of
CC   Helicobacter in a sample. The products are also of use in screening
CC   for compounds having the ability to interfere with the H. pylori life
CC   cycle or to inhibit H. pylori infection.
SQ   Sequence   282 AA;

  Query Match           11.0%;  Score 176;  DB 1;  Length 282;
  Best Local Similarity  25.1%;  Pred. No. 5.19e-04;
  Matches    56;  Conservative   59;  Mismatches  90;  Indels  18;  Gaps  14;

Db      64 PLDFNY-PV-HIVQAPQNHHVVGILMPRIQVSDNL-KPYIDKFQDALINQIQTIFEKRGY 120
           :| :|| |    |||  :  :: :| | :| |||: | | :|| :     :: |::::||
Qy      39 ALKLNYHPASEKVQA-LDEKIL-LLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db     121 QVLRFQD-EKA-LNVQDKKKIFSVLDLKGWVGILEDLKMNL--K-DPN---SPNLDTL-- 170
            |:  :  :|  ::  :||  : : ::| : :  | |  :  | :|    | ||  :
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     171 VDQSSGSVWFNFYEPESNRVVHDFAVEVGTFQAITYTYTSTNNASGGFNSSKSVIHENLD 230
           |  ::| :   : || |   :  |::::: :: |   :  | ::| :     :::    |
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSELD-IQEKFLKTTHSSHSGGLVSTMVKGT-D 214

Db     231 KNREDAIHKILNRMYAVVMKKAVTELTKENIAKYRDAIDRMKG 273
           | :||| : ||:::| :|      ||  |:  |:     :||
Qy     215 -NSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKG 256
```

```
RESULT   14
ID   W20938 standard; protein; 285 AA.
AC   W20938;
DT   21-JUL-1997  (first entry)
DE   H. pylori secreted or periplasmic protein, 27ze10351orf5.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS   Helicobacter pylori.
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   N-PSDB; T68191.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Claim 72; Page 1333-1334; 1481pp; English.
CC   The present sequence is a H. pylori secreted or periplasmic protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
SQ   Sequence   285 AA;

  Query Match           11.0%;  Score 176;  DB 1;  Length 285;
  Best Local Similarity  25.1%;  Pred. No. 5.19e-04;
  Matches    56;  Conservative   59;  Mismatches  90;  Indels  18;  Gaps  14;

Db      67 PLDFNY-PV-HIVQAPQNHHVVGILMPRIQVSDNL-KPYIDKFQDALINQIQTIFEKRGY 123
           :| :|| |    |||  :  :: :| | :| |||: | | :|| :     :: |::::||
Qy      39 ALKLNYHPASEKVQA-LDEKIL-LLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db     124 QVLRFQD-EKA-LNVQDKKKIFSVLDLKGWVGILEDLKMNL--K-DPN---SPNLDTL-- 173
            |:  :  :|  ::  :||  :  : ::| : :  | |  :  | :|    |  || :
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNGEIVLRPDPKRTIQKKSEPGLLFSTGLDKMEG 156

Db     174 VDQSSGSVWFNFYEPESNRVVHDFAVEVGTFQAITYTYTSTNNASGGFNSSKSVIHENLD 233
           |  ::| :   : || |   :  |::::: :: |   :  | ::| :     :::    |
Qy     157 VLIPAGFIKVTILEPMSGESLDSFTMDLSELD-IQEKFLKTTHSSHSGGLVSTMVKGT-D 214

Db     234 KNREDAIHKILNRMYAVVMKKAVTELTKENIAKYRDAIDRMKG 276
            | :||| : ||:::| :|      ||  |:  |:     :||
Qy     215 -NSNDAIKRALNKIFANIMQEIDKKLTQKNLESYQKDAKELKG 256


RESULT   15
ID   W20468 standard; Protein; 147 AA.
AC   W20468;
DT   11-JUL-1997  (first entry)
DE   H. pylori secreted or peripplsmic protein, 3906963.aa.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacterium; life cycle; activator;
KW   bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW   diagnosis.
OS   Helicobacter pylori.
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   N-PSDB; T67633.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Claim 72; Page 638; 1481pp; English.
CC   This sequence represents a H. pylori secreted or periplasmic protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
SQ   Sequence   147 AA;

  Query Match           7.4%;  Score 118;  DB 1;  Length 147;
  Best Local Similarity  28.1%;  Pred. No. 3.41e+00;
  Matches    25;  Conservative   26;  Mismatches  31;  Indels   7;  Gaps   7;

Db      64 PLDFNY-PV-HIVQAPQNHHVVGILMPRIQVSDNL-KPYIDKFQDALINQIQTIFEKRGY 120
           :| :|| |    |||  :  :: :| | :| |||: | | :|| :     :: |::::||
Qy      39 ALKLNYHPASEKVQA-LDEKIL-LLRPAFQYSDNIAKEYENKFKNQTALKVEQILQNQGY 96

Db     121 QVLRFQD-EKA-LNVQDKKKIFSVLDLKG 147
            |:  :  :|  ::  :||  :  : ::|
Qy      97 KVISVDSSDKDDFSFAQKKEGYLAVAMNG 125


Search completed: Tue Sep 26 16:19:09 2000
Job time : 12 secs.
```